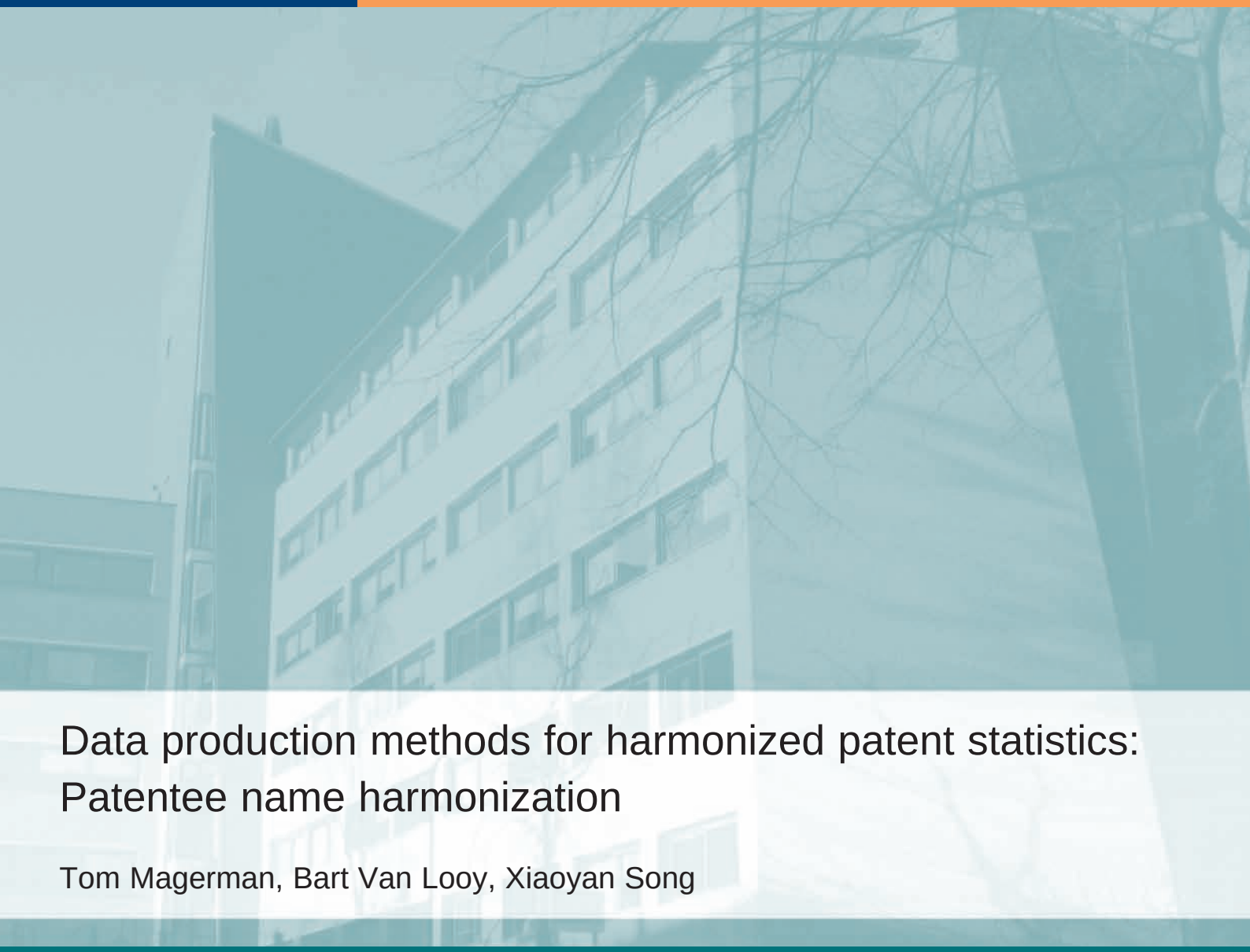
KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Economics and Applied Economics



# Data production methods for harmonized patent statistics: Patentee name harmonization

Tom Magerman, Bart Van Looy, Xiaoyan Song

DEPARTMENT OF MANAGERIAL ECONOMICS, STRATEGY AND INNOVATION (MSI)

MSI 0605

# Data production methods for harmonized patent statistics: Patentee name harmonization


Tom Magerman[2, 3], Bart Van Looy[1, 2, 3], Xiaoyan Song[3]

[1]*Managerial Economics, Strategy and Innovation, Faculty of Ecomomics & Applied Economics, K.U.Leuven, Naamsestraat 69, B-3000 Leuven, Belgium*

[2]*Research Division INCENTIM, Faculty of Economics & Applied Economics, K.U.Leuven, Naamsestraat 69, B-3000 Leuven, Belgium*

[3]*Steunpunt O&O Statistieken, Dekenstraat 2, B-3000 Leuven, Belgium*

Leuven, March 2006




*Corresponding author E-mail address: tom.magerman@econ.kuleuven.be*

*Recommended citation: Magerman T., Van Looy B., Song X (2006), "Data production methods for harmonized patent statistics: Patentee name harmonization", K.U.Leuven FETEW MSI Research report 0605, Leuven*

# 1 INTRODUCTION

Patent documents are one of the most comprehensive data sources on technology performance. Although technology indicators based on patent documents have certain limitations[1], Griliches' observation of almost two decades ago still seems to hold: "*In spite of all the difficulties, patent statistics remain a unique resource for the analysis of the process of technical change. Nothing else even comes close in the quantity of available data, accessibility, and the potential industrial, organizational and technological detail.*" (Griliches, 1990). Patent indicators are now used by companies and by policy and government agencies[2] alike to assess technological progress on the level of regions, countries, domains[3], and even specific entities such as companies, universities and individual inventors. However, with respect to the latter (i.e. analysis on the level of the patentee), specific concerns can be discerned.

These concerns stem from the heterogeneity of patentee names to be found in patent documents. The same organization or individual can appear in different guises when patentees apply for patents through different channels over extended time periods. While this poses no specific challenge to the functioning of the patent system itself – where patent documents are used on a recurrent basis to assess prior art – it complicates the analysis on the level of patentees. The analyst is confronted with inconsistencies such as spelling mistakes, typographical errors and name variants, which often reflect idiosyncrasies in the organization of research and intellectual property right activity at particular moments within one and the same organization.

These discrepancies in the naming of identical patentees in current patent databases justify efforts to achieve name harmonization so that analysis at the level of patentees can be facilitated. Quality, in terms of both completeness and accuracy, is a crucial issue in this respect. We refer to 'completeness' as the extent to which the name-harmonization procedure is able to capture all name variants of the same patentee. 'Accuracy' relates to the extent to which the name-harmonization procedure correctly allocates name variants to a single, harmonized patentee name. Unfortunately, completeness and accuracy do not go hand in hand. Efforts directed to maximizing the number of identified name variants will ultimately lead to decreasing accuracy, while maximizing accuracy inevitably leads to an increase in missed or unidentified name variants, or decreasing completeness.

In this paper, we develop a comprehensive method to achieve harmonization of patentee names in an automated way. The method has been applied to an extensive set of all patentee names found for all EPO patent applications published between 1978 and 2004 and all granted USPTO patents published between 1991 and 2003. Priority has been given to accuracy, as demonstrated in section 4 – A content-driven name harmonization approach focusing on accuracy.

Before discussing in detail the methodology and its effects as applied to the EPO and USPTO patentee name list, we will first clarify the difference between patentee name harmonization and legal entity identification. In addition, we will briefly expand on the methods and approaches previously developed to address the issue of patentee name harmonization, in order to shed light on our specific contribution. Finally, future refinements and extensions are discussed.

---

[1] Propensities to patent differ among industries, firms and countries.

[2] Patent indicators are now to be found in recurrent publications of the National Science Foundation (US), the European Commission (Science and Technology Indicator Reports) and the OECD alike.

[3] Analysis by domains is feasible by using the WIPO International Patent Classification or aggregation schemas like the 'Systematic of OST/INPI/FhG ISI of 5 technology areas and 30 sub-areas'; analysis in relation to industries is enabled by concordance schemes based on patent classification, like the MERIT concordance table (Verspagen, 1994), the OECD Technology Concordance (Johnson, 2002), or the EC DG Research and FhG ISI/OST/SPRU concordance table (Schmoch, Laville, Patel, Frietsch, 2003).

# 2 PATENTEE NAME HARMONIZATION AND LEGAL ENTITY HARMONIZATION

The focus of the methodology outlined in this paper is on patentee name harmonization. This does not equate to harmonization on the level of the legal entity. Legal entity harmonization is concerned with the identification of all patents owned by one and the same legal entity. In this respect, legal entity harmonization is not only concerned with name inconsistencies but takes mergers and acquisitions, name changes, and subsidiaries into account. For instance, when aiming at legal entity harmonization, all patents held by Hewlett Packard, Digital Equipment Corporation and Compaq might be considered as belonging to one and the same legal entity; likewise, "ANDERSEN CONSULTING" would become harmonized to "ACCENTURE" (name change).

In other words, when harmonizing legal entities, every patentee name needs to be checked against historical information on naming practices and ownership in order to address the following issues:

- Identification of entities (business units, departments, subsidiaries) that may have a different name but belong to the same legal entity;
- Identification of name changes over time;
- Identification of mergers and acquisitions;
- Identification of joint ventures;
- Identification of mother and daughter relationships / subsidiary companies.

It is clear that this level of information is not available in current patent databases. External information is needed - on ownership, changes of ownership, and organizational practices with regard to names - to arrive at a comprehensive methodology for legal entity harmonization. Given the absence of databases providing exhaustive coverage of information needed to achieve legal entity harmonization[4], such efforts are not included in the name-harmonization methodology outlined in this paper.

Accordingly, our methodology focuses on the identification of name variations by comparing each patentee name with all other patentee names; the objective is to match names that appear to be similar but differ because of spelling or language variations. The same patentee name can appear in a different form in the patentee name list for the following reasons:

- Spelling variations (different but correct spelling variations), e.g. "IBM" and "I.B.M.", or "BAIN & CO" and "BAIN AND COMPANY";
- Typographical errors, e.g. "INTERNATIONAL BUSINESS MACHINES" and "INTERATIONAL BUSINESS MACHINES";
- Addition of the legal form (again with possible acronyms, spelling variations, mistakes, and typographical errors in the legal form), e.g. "IBM", "IBM CORP.", "IBM CORPORATION" and "IBM COPRORATION", or "BAYER", "BAYER A.G." and "BAYER AG";
- Errors, e.g. "INTERNATIONAL BUSINESS MACHINES" and "INTELLIGENT BUSINESS MACHINES";
- Addition of establishment, business unit, department, subsidiary name or geographic identifier, e.g. "IBM" and "IBM JAPAN";
- Acronyms, e.g. "IBM" and "INTERNATIONAL BUSINESS MACHINES".

All of these issues will be analyzed in a systematic manner in order to develop an appropriate methodology. It will become apparent that spelling variations, typographical errors and the additions of legal forms can be addressed in an automated manner while for errors, acronyms and business unit or department extensions additional validation efforts will be

[4] While information providers like Graydon, Dunn & Bradstreet, Bureau Van Dijk and Thomson Scientific offer data on mergers and acquisitions and subsidiaries, this information is limited to larger entities and/or is confined to more recent years.

required in order to be accurate. However, before discussing in detail the methods and their impact in detail, it can be noted that name harmonization efforts concerning patentee names have been undertaken in the past, notably by USPTO and by Derwent (Thomson Scientific).,Before discussing the development of the name cleaning and harmonization procedures proposed in this paper, these approaches will be first briefly discussed.

# 3 EXISTING NAME HARMONIZATION APPROACHES

## 3.1 USPTO CONAME assignee name harmonization

As part of the USPTO TAF database, first-named assignee names of organizational entities are harmonized for utility patents granted since 1969.

The USPTO harmonization rules are conservative, as further consolidation of names is considered far easier then separating combined names. Harmonization efforts do not address subsidiary ownership but are limited to identify assignee name variations. In addition, organizations with similar names but associated with different countries or a different legal form are not harmonized.

In the case of patents granted prior to July 1992, harmonization is primarily based on a manual process of comparing names. For patents granted after July 1992, harmonization is largely based on an automated procedure. This procedure can be summarized as follows:

- Extract name of first-named assignee;
- Condense assignee name by removing spaces and non-alphanumeric characters;
- Convert to uppercase characters;
- Match condensed name with existing list of condensed and harmonized names;
- Manual review all new assignee names not yet matched to an existing name in previous step (e.g. by looking at assignees of other patents granted to the same inventor or inventors);
- Annual large scale manual review to verify integrity of the entire assignee file.

The partial manual approach of USPTO offers potential to achieve high levels of completeness. Especially the 'staging' approach, whereby new names not yet matched are compared with previously harmonized names, allows for a complete harmonization solution.

The USPTO harmonization has however following shortcomings:

- The partial manual approach implies significant resources every time new patentee names appear in the database;
- Only the first assignee is processed;
- Names reflecting different legal forms or associated with different countries are not combined[5];
- The manual review process is not transparent and might cause rule variation since harmonization is performed by different persons, jeopardizing the reproduction on a broader set of names (e.g. EPO applicant names, second assignee) [6].

## 3.2 DERWENT WPI company name harmonization

The DERWENT WORLD PATENT INDEX provides patentee codes for all patentees. One can summarize the DERWENT WPI method to produce these patentee codes as follows[7]:

---

[5] For example, in the USPTO harmonization, the following name variations of "BURR-BROWN" can be found in the list of harmonized names: "BURR-BROWN CORPORATION", "BURR-BROWN INC." and "BURR-BROWN LIMITED".

[6] For instance, this can be observed in the list of original assignee names harmonized to "AT&T CORP.": "Bell Telephone Laboratories Inc.", "AT&T Corp/CSI Zeinet (A Cabletron Co.)", "ATT Corp--Lucent Technologies Inc" and "AT&T Middletown". It is clear that some of these names are associated with "AT&T Corp." based on criteria other than name similarity. However, it remains unclear which additional rules have been applied and to what extent.

[7] For a more detailed description, see: http://www.thomsonscientific.com/media/scpdf/patenteecodes.pdf

- Take the name and replace commonly occurring words with a standardized version or abbreviation, as listed in the DERWENT abbreviated word list (Russian and Japanese words are first translated to English);
- Select the first significant word(s) of the resulting name, ignoring 'common' words listed in the DERWENT list of common descriptors;
- Replace frequently occurring words recorded in the DERWENT list of general descriptors with a two-letter abbreviation;
- Replace continent, country, region and town names with a two-letter abbreviation (some commonly used names are replaced with three-letter abbreviations);
- Replace points of the compass with one- or two-letter abbreviations;
- Take the first four letters of the remaining word.

This results in a long list of so called non-standard patentee codes consisting of four letters. These codes are not necessarily unique; several unrelated patentees can have the same automatically generated patentee code[8].

Next, a selection of these patentees is analyzed in depth to arrive at unique standard patentee codes. Within this phase the emphasis shifts towards legal entity harmonization. This latter objective is achieved by incorporating additional information on companies derived from secondary financial sources. These efforts are however limited to patentees applying for larger numbers of patent applications. This reduction is understandable since arriving at standard patentee codes in the WPI approach implies legal entity harmonization: mergers and acquisitions, name changes and subsidiaries.

At present, the index of standard patentee codes provided by WPI contains 21,000 entities and can be considered the most comprehensive harmonized index currently available since it includes legal entity harmonization. At the same time, the process to arrive at standard names is not transparent and case specific (for example, standard codes are retained for company name changes, but in case of mergers and acquisitions, either one of the codes is retained and the others abandoned, either a new code is created). The precise rules that have been applied in each case are only evident after the names associated with a certain standard patentee code have been analyzed (information which is not publicly available)[9].

For companies for which a standard code is not available (because having only a limited number of patents), or not recognizable as a subsidiary of a company already having a standard code, the automatically generated non-standard code cannot be considered appropriate to achieve harmonization of the complete list of patentee names. The rules to come to the non-standard code result in numerous false matches and low level of accuracy[10].

[8] For example, the non-standard code "HUSS" is associated with "HUSSMANN CORP", "HUSSOR SA", "HUSSOR ERECTA SA", "HUSS MASCHFAB GMBH & CO KG", "HUSS UMWELTTECHNIK GMBH" and "HUSSMANN DO BRASIL LTDA".

[9] For example, the standard code "CANO" is associated with "CANON CAMERA", "CANON KK", "CANON PRECISION INC", "CANON PRECISION MAC" and "CANON SEIKI KK". Another standard code "CAND" is associated with "CANON DENSHI KK", "CANON ELECTRONICS CO LTD" and "CANON ELECTRONICS INC".

[10] These non-standard codes are however useful because they provide a high level of completeness, resulting in a maximum set of names that might be combined.

# 4 A CONTENT-DRIVEN NAME HARMONIZATION APPROACH FOCUSING ON ACCURACY

As indicated in the introduction, name harmonization involves a trade-off between completeness and accuracy. It has been a deliberate choice in the methodology outlined here to favor accuracy over completeness for reasons of transparency, as it is easier to combine additional names than separate combined names. An accurate but somewhat incomplete set of harmonized names provides users with ample opportunities to extend the methodology and its results to a broad range of applications. Given an accurate set of harmonized names, additional name matches that are considered relevant can be identified and added in a straightforward manner. Reverse operations, starting with a more complete set, are much more complicated since previous steps undertaken to achieve a more complete result might need to be undone or 'reverse engineered'. In practice, this would prove to be a much more complicated endeavor than combining disaggregated names. Hence, this methodology, conceived as a transparent and accurate set of harmonized names in which completeness can be gradually improved, is considered far more appealing than a more complete set which contains the risk of not being accurate or being unsuited to specific analytical purposes.

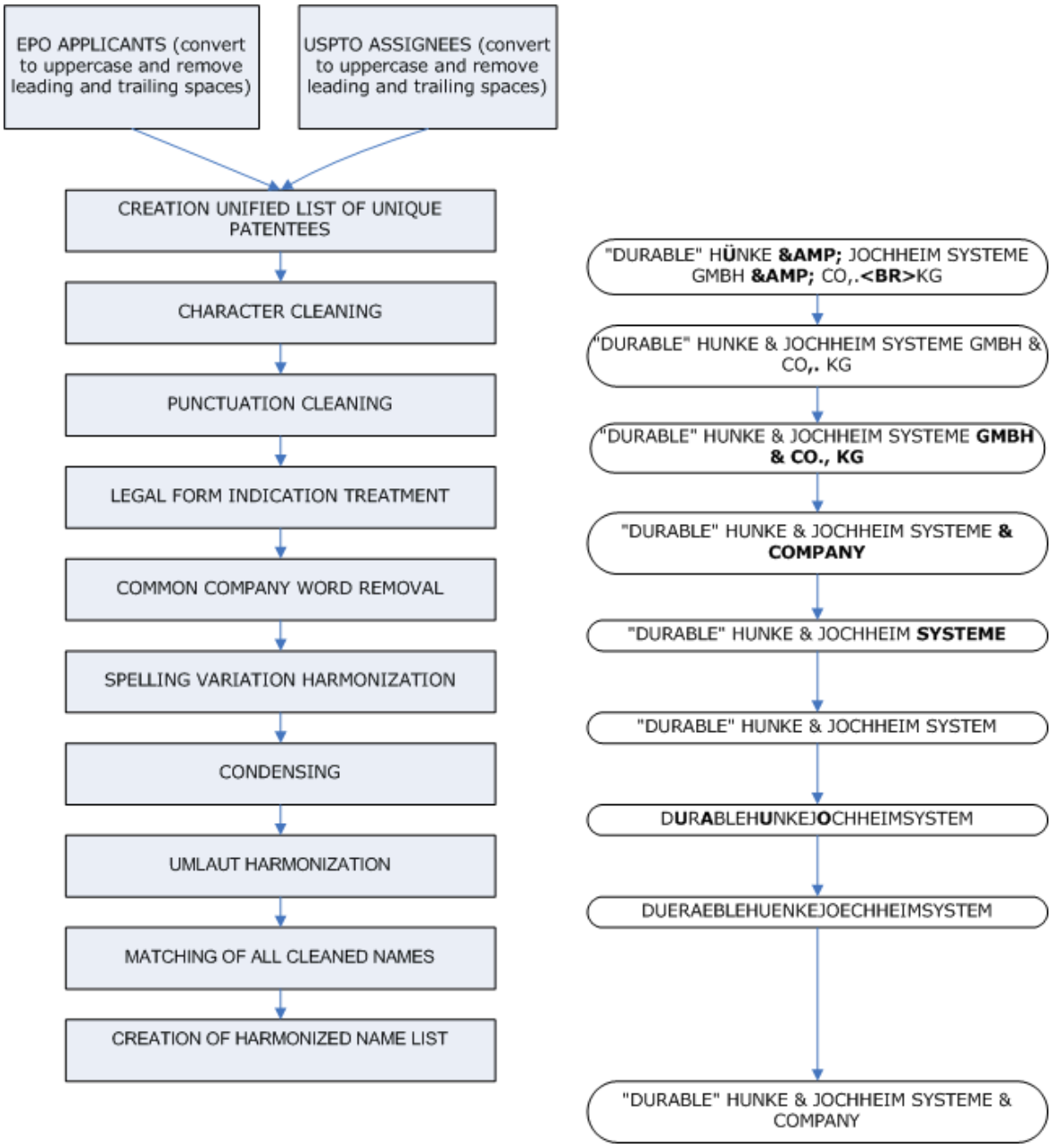
As a result, the development of the methodology is based on the underlying principle that every step in the cleaning and harmonization process must increase completeness without decreasing accuracy. Every action that jeopardizes accuracy will ultimately be excluded from the methodology, as combining two names belonging to two different legal entities has to be avoided at all cost. Moreover, in order to achieve sufficient levels of accuracy, several of the procedures and rules that have been developed take into account the specificities of the full original name list. This content-driven approach results in a partly manual, and hence labor-intensive, development process.

The final procedure can be completely automated in a modular approach to allow further refinements and improvements. The entire procedure is organized as a series of generic steps and sub-steps that are implemented by taking into account the nature of the source data. It should be noted that while the more generic parts of the procedure can be used for all kinds of name-harmonization applications, some procedures are highly content-specific and additional analysis and refinements might be needed if applied to a different set of organization names.

Figure 1 contains an overview of the comprehensive methodology that consists of a sequence of steps, including both data pre-processing and name-harmonizing activities. An example patentee name is included to see the results of each step (string parts that will be affected in the next processing step are highlighted in bold).

Appendix 1 describes in detail all steps in the name-cleaning and harmonization procedure as applied to the dataset with EPO applicants and USPTO assignees, including examples, detailed analysis and implementation. The main principles underlying each step are explained in the following paragraphs. This will facilitate discussion of the results in section 5 – Results and Impact.

**Figure 1: Overview schema name cleaning and harmonization**




## 4.1 Data pre-processing

In the pre-processing steps, data are prepared for processing to facilitate actual name cleaning and harmonization. The individual impact of each step on the number of unique patentee names is limited but it smoothes progression through consecutive steps and considerably increases the overall impact. Data pre-processing is highly dependent on the content of the underlying data. Consequently, extensive refinements or adaptations may be needed when processing names from a different data source.

### 4.1.1 Character cleaning

Depending on the data source, non-letter (A to Z) and non-digit (0 to 9) characters can be coded or represented in a variety of ways (e.g. ANSI, SGML), inducing additional name

variations. Data can also contain codes that bear no relation to the real data and merely represent formatting issues; again, this induces additional name variations.

Character cleaning removes different types of character representations and formatting codes or converts them to genuine standard ASCII characters. For instance, HTML formatting codes such as "<BR>" are removed or replaced by spaces and SGML codes such as "&OACUTE;" are removed or replaced by their ASCII equivalent whenever possible.

In this step, names are also scanned for proprietary coded characters like "{UMLAUT OVER (A)}" in USPTO data. These codes are also removed or replaced whenever possible. Accented characters like "É" are replaced with their unaccented ASCII equivalents. Particular problems with alternative spellings of the umlaut in German (and some other languages) are treated at a later stage (see section 4.2.5 – Umlaut harmonization).

### 4.1.2 Punctuation cleaning (pre-parsing)

Names may not only contain letters and digits but also characters such as ",", ";", and "-", used to separate words or to indicate abbreviations and combinations. These characters might complicate the separation or parsing of names into individual words, which is necessary in further cleaning steps (e.g. identifying the legal form). Punctuation cleaning aims to harmonize all of these punctuation characters, and, thereby, facilitate the parsing of names in individual words at a later stage.

Firstly, double spaces are replaced with single spaces. Quotation marks followed by a space appearing at the beginning of a name, or preceded by a space appearing at the end of a name, are replaced with quotation marks without a trailing or leading space. Quotation marks are removed from names having only quotation marks at the beginning and the end of the name. Next, names are scanned for non-alphanumerical characters at the beginning and the end of the name, and these characters are removed if appropriate. Finally, comma and period irregularities are harmonized so that commas are not preceded by spaces but followed by a space (unless acting as decimal or thousand separators) and periods are only preceded by letters or digits.

## 4.2 Name cleaning

In the name cleaning steps, the actual name cleaning and harmonization is performed. As mentioned previously, our approach takes the content of the data into account; extensive refinements or adaptations might be needed when names from a different data source are processed.

### 4.2.1 Legal form indication treatment

A lot of patentee names contain some kind of legal form indication (e.g. "INC.", "LIMITED", and "LTD."). These legal form indications are responsible for a considerable number of name variations due to the variety of abbreviations and spellings used. In this step, legal form indications are harmonized and moved to a separate field, thereby considerably reducing name variations.

### 4.2.2 Common company word removal

Legal form indications are separated out since they do not constitute a distinctive part of the name; this logic applies to some other words as well. In the case of companies especially, additional words like "COMPANY", "CORPORATION", "GESELLSHAFT" and "SOCIETE" add nothing to the distinctive character of a company name. When two names are found to be identical except for the presence of such words, the underlying patentee name will be taken as referring to one and the same organization. Examples include "3COM" and "3COM CORPORATION", "AMIC" and "AMIC COMPANY", "BAUR SPEZIALTIEFBAU" and "BAUR SPEZIALTIEFBAU GESELLSCHAFT", and "SOCIETE NOVATEC" and "NOVATEC".

### 4.2.3 Spelling variation harmonization

Typographical errors and spelling mistakes are responsible for considerable name variations. These kinds of error can be identified by assessing word similarities. Whilst this type of analysis is straightforward for common English words, proper names usually require manual validation efforts in order to ensure accuracy. For example, "AMTECH" and "IMTECH" only differ in a single character but it would be incorrect to automatically assume that the names refer to one and the same patentee. For common words, spelling and language variations can be identified without ambiguity and, therefore, harmonized effortlessly. For example, "SYSTEM", "SYSTEMS", "SYSTEMEN", and "SYSTEMES" can all be harmonized to "SYSTEM" or "SYSTEMS". Spelling variation harmonization replaces all variants of common words with one harmonized variant that will be used to match name variants.

### 4.2.4 Condensing

Significant name variations are also caused by word separation, punctuation, and non-alphanumerical characters, which clearly have no relevance in identifying the distinctive characteristics of a name (e.g. "3 COM" and "3COM", and "AAF-MCQUAY", "AAF MCQAY" and "AAF – MCQAY"). Condensing removes all non-alphanumerical characters so that a harmonized variant can be used to match names.

### 4.2.5 Umlaut harmonization

Although accented characters have already been replaced (see section 4.1.1 – Character cleaning), German characters with a diacritic mark (umlaut: "ä", "ö", "ü") still generate spelling variations because words containing them can occur in three varieties, one with an umlaut (e.g. "für"), an alternative spelling without an umlaut but with an additional "e" (e.g. "fuer"), and a simplified form without both an umlaut and an additional "e" (e.g. "fur"). Umlaut harmonization identifies and matches different variants of words including "ä", "ö" and "ü".

# 5 RESULTS AND IMPACT

The complete name cleaning and harmonization procedure has been applied to an integrated set of EPO and USPTO patentee names. In the case of the EPO dataset, all 270,635 applicant names are included from all 1,600,812 EPO patent applications published between 1978 and 2004 (based on the EPO ESPACE ACCESS product). For USPTO data, all 223,665 assignee names are included from all 1,614,224 USPTO granted patents published between 1991 and 2003 (based on the USPTO Grant Red Book product). All names were converted to uppercase. Combining these two datasets produced a name list of 443,722 unique patentee names (after conversion to uppercase and removal of leading and trailing spaces). Of these names, 50,578 appear both in EPO and USPTO; 173,087 only appear in USPTO, while 220,057 only appear in EPO.

As indicated previously, Appendix 1 contains a detailed description of the algorithms that have been applied and a detailed analysis of the impact of the name cleaning and harmonization procedure on the set of EPO and USPTO patentee names. In this section, we highlight the major findings.

Overall, harmonization has reduced the number of unique patentee names by 17.6%, from 443,722 to 365,866 names. The average number of patents per patentee increases from 7.2 before to 8.8, after harmonization.

13.5% or 49,449 of the harmonized names are matched with more than one original patentee name. The distribution of the number of matched names is skewed, ranging from 2 to 51 original names, with an average of 2.6.

When names are matched, an average increase in patent volume of 7.2 patents is observed. The most extreme case is IBM for which name harmonization results in an additional 12,704 patents. When comparing the number of additional patents allocated to harmonized names with the highest volume pertaining to one of the matched set of names, an average increase of 33.1% is observed. The most extreme case in this respect produces an increase of 86% ("H. G. WEBER & COMPANY").

The number of patentees with a low number of patents decreased considerably but remains large; the number of patentees having only one patent dropped by 22% from 281,760 to 219,821 (60.13%).

The top 10 patenting organizations are identical except for "INTERNATIONAL BUSINESS MACHINES CORPORATION" moving from third to first place. Greater variation is observed when considering the Top 25. The lower the absolute numbers become, the more the rankings will be impacted. Table 1 contains the top 25 patentees after name cleaning and harmonization, with the ranking and the number of patents before and after harmonization.

**Table 1: Ranking of patentees before and after harmonization**

| RANK AFTER | RANK BEFORE | DIFF | HARMONIZED NAME | PAT AFTER | PAT BEFORE | DIFF |
|---|---|---|---|---|---|---|
| 1 | 3 | +2 | INTERNATIONAL BUSINESS MACHINES CORPORATION | 41,173 | 28,469 | 12,704 |
| 2 | 1 | -1 | CANON | 31,741 | 31,649 | 92 |
| 3 | 2 | -1 | SIEMENS | 30,770 | 30,452 | 318 |
| 4 | 4 | - | MATSUSHITA ELECTRIC INDUSTRIAL COMPANY | 26,379 | 25,594 | 785 |
| 5 | 5 | - | SONY CORPORATION | 23,665 | 23,620 | 45 |
| 6 | 6 | - | NEC CORPORATION | 23,508 | 23,468 | 40 |
| 7 | 7 | - | TOSHIBA | 23,344 | 23,277 | 67 |
| 8 | 8 | - | HITACHI | 22,754 | 22,226 | 528 |
| 9 | 9 | - | GENERAL ELECTRIC COMPANY | 19,620 | 19,117 | 503 |
| 10 | 10 | - | EASTMAN KODAK COMPANY | 18,863 | 18,847 | 16 |
| 11 | 12 | +1 | FUJITSU | 18,575 | 18,310 | 265 |
| 12 | 11 | -1 | MITSUBISHI DENKI | 18,513 | 18,445 | 68 |
| 13 | 14 | +1 | BASF | 18,499 | 16,855 | 1,644 |
| 14 | 15 | +1 | MOTOROLA | 17,294 | 15,758 | 1,536 |
| 15 | 24 | +9 | BAYER | 17,220 | 10,053 | 7,167 |
| 16 | 13 | -3 | ROBERT BOSCH | 17,052 | 16,870 | 182 |
| 17 | 17 | - | SAMSUNG ELECTRONICS COMPANY | 14,897 | 13,561 | 1,336 |
| 18 | 16 | -2 | KONINKLIJKE PHILIPS ELECTRONICS | 14,550 | 14,411 | 139 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 19 | 18 | -1 | FUJI PHOTO FILM COMPANY | 12,985 | 12,652 | 333 |
| 20 | 36 | +16 | E.I. DU PONT DE NEMOURS & COMPANY | 12,252 | 6,269 | 5,983 |
| 21 | 19 | -2 | XEROX CORPORATION | 12,111 | 12,104 | 7 |
| 22 | 21 | -1 | HEWLETT-PACKARD COMPANY | 12,024 | 11,018 | 1,006 |
| 23 | 32 | +9 | THE PROCTER & GAMBLE COMPANY | 11,862 | 7,300 | 4,562 |
| 24 | 22 | -2 | SHARP | 10,880 | 10,584 | 296 |
| 25 | 23 | -2 | TEXAS INSTRUMENTS | 10,801 | 10,353 | 448 |

To sum up, the impact of this harmonization effort is considerable. At the same time, the impact on specific patentees varies greatly; while some patentees gain significantly in terms of the absolute (up to 12,704) or relative (up to 86%) number of patents, no changes are observed for the majority of patentees. Rankings will be affected, especially when one considers sub-samples characterized by a smaller number of patents.

Of course, the methodology as outlined can be extended. In this respect, algorithms such as approximate string searching or approaches oriented towards identifying acronyms seem to be highly relevant. Furthermore, introducing address field information might result in additional improvements. In the following section, we will outline these approaches and their relevancy for name harmonization. At the same time, it will become apparent that these extensions cannot be implemented in a fully automated way. Since they imply validation efforts of a manual nature, they have not been integrated into the comprehensive methodology detailed in Appendix 1.

# 6 DIRECTIONS FOR FURTHER DEVELOPMENT

## 6.1 Approximate string searching

Approximate string searching aims to identify patentee names that are slightly different but might, in fact, be the same. Rather than applying exact string matching, distance measures are used based on the similarity between the strings, However, it will become apparent that one is confronted with a trade-off between accuracy and completeness: identifying similar patentee names might result in matching unrelated patentee names and vice versa.

Several algorithms are available (bigrams, trigrams) but using the Levenshtein measure (edit-distance measure), which computes string similarity, can be considered an appropriate and easy-to-implement starting point (see for an example of an implementation approach, the AGREP tool[11]). When using the Levenshtein distance, a string $P$ is said to be at distance $k$ to a string $Q$ if $P$ can be transformed to equal $Q$ with a sequence of $k$ insertions of single characters in (arbitrary places in) $P$, deletions of single characters in $P$, or substitutions of characters[12].

Using this measure, all patentee names within a certain distance of a given patentee name can be identified.

However, implementation is not straightforward because of the following issues:

- Performance: approximate string searching is not as fast as exact string searching. Moreover, if a list has to be matched with itself (e.g. one wants to compare all names in a list with all other names in that list), processing time will increase exponentially with the number of items in the list (a list of 100 items requires 10,000 distances to be calculated and assessed; a list of 1,000 items requires 1,000,000 calculations);
- Large output: approximate string searching can produce a long list of matched strings. Moreover, if a list is matched with itself, numerous duplicates will appear in the matching list (if a string matches 50 other strings, the same combinations will appear 50 x 50 times).
- Relevance of error: if the Levenshtein distance is used, the acceptable distance is dependent on the length of the string. For a string of 3 characters, name variants with a Levenshtein distance of 1 will in most cases result in false hits (e.g. NEC, NAC, NIC, NEI, and NEA). A minimum string length might be required.
- Validation: even if a small Levenshtein distance between two larger strings is observed, this does not automatically mean that the underlying names are identical. A difference of one character can result in significantly different names. In order to be accurate, additional validation efforts need to be undertaken.

In view of these issues, applying approximate string searching in an automated way for all names within the name list has not been included in the outlined method. At the same time, approximate string searching offers significant potential especially for longer names, as the following illustrations demonstrate. Implementing such efforts does however imply extensive – manual – validation efforts before they can become automated.

Table 2 contains the name variations of "INTERNATIONAL BUSINESS MACHINES" identified with approximate string searching (legal form indications were removed first and resulting names were condensed).

**Table 2: "INTERNATIONAL BUSINESS MACHINES" name variations identified by approximate string searching**

| |
|---|
| IBMCORPINTERNATIONALBUSINESSMACHINESCORP |
| IBMCORPORATIONINTERNATIONALBUSINESSMACHINESCORPORATION |

[11] S. Wu and U. Manber, "AGREP -- A fast approximate pattern-matching tool", proc. Winter 1992 USENIX Technical Conference

[12] Variants of this logic introduce weights for the different operations; e.g. deletions 3, insertions 2, and substitutions 1.

IBMINTERNATIONALBUSINESSMACHINES
IINTERNATIONALBUSINESSMACHINES
INERNATIONALBUSINESSMACHINES
INTENATIONALBUSINESSMACHINES
INTERANATIONALBUSINESSMACHINES
INTERNAIONALBUSINESSMACHINES
INTERNATIAONALBUSINESSMACHINES
INTERNATIIONALBUSINESSMACHINES
INTERNATINALBUSINESSMACHINES
INTERNATIOALBUSINESSMACHINES
INTERNATIOANALBUSINESSMACHINES
INTERNATIOINALBUSINESSMACHINES
INTERNATIONAALBUSINESSMACHINES
INTERNATIONALBUINESSMACHINES
INTERNATIONALBUNSINESSMACHINES
INTERNATIONALBUSINBESSMACHINES
INTERNATIONALBUSINEESMACHINES
INTERNATIONALBUSINEMSSMACHINES
INTERNATIONALBUSINESMACHINES
INTERNATIONALBUSINESSMACHINCES
INTERNATIONALBUSINESSMACHINE
INTERNATIONALBUSINESSMACHINECORPINTERNATIONALPROPERTYLAW
INTERNATIONALBUSINESSMACHINES
INTERNATIONALBUSINESSMACHINESC
INTERNATIONALBUSINESSMACHINESCIRPORATION
INTERNATIONALBUSINESSMACHINESCOIRPORATION
INTERNATIONALBUSINESSMACHINESCOMPANY
INTERNATIONALBUSINESSMACHINESCOPORATION
INTERNATIONALBUSINESSMACHINESCOPROATION
INTERNATIONALBUSINESSMACHINESCOPRORATION
INTERNATIONALBUSINESSMACHINESCORORATION
INTERNATIONALBUSINESSMACHINESCORP
INTERNATIONALBUSINESSMACHINESCORPOATION
INTERNATIONALBUSINESSMACHINESCORPORAITON
INTERNATIONALBUSINESSMACHINESCORPORARTION
INTERNATIONALBUSINESSMACHINESCORPORATAION
INTERNATIONALBUSINESSMACHINESCORPORATIION
INTERNATIONALBUSINESSMACHINESCORPORATIN
INTERNATIONALBUSINESSMACHINESCORPORATIOIN
INTERNATIONALBUSINESSMACHINESCORPORATIOM
INTERNATIONALBUSINESSMACHINESCORPORATION
INTERNATIONALBUSINESSMACHINESCORPORATIONIBM
INTERNATIONALBUSINESSMACHINESCORPORATIONS
INTERNATIONALBUSINESSMACHINESCORPORATOIN
INTERNATIONALBUSINESSMACHINESCORPORATON
INTERNATIONALBUSINESSMACHINESCORPORATRION
INTERNATIONALBUSINESSMACHINESCORPORTAION
INTERNATIONALBUSINESSMACHINESCORPORTATION
INTERNATIONALBUSINESSMACHINESCORPORTION
INTERNATIONALBUSINESSMACHINESCORPROATION
INTERNATIONALBUSINESSMACHINESINCORPORATION
INTERNATIONALBUSINESSMACHINESMACHINE
INTERNATIONALBUSINESSMACHINESMACHINES
INTERNATIONALBUSINESSMACHINESOPERATION
INTERNATIONALBUSINESSMACHINESS
INTERNATIONALBUSINESSMACHNES
INTERNATIONALBUSINESSMACHNINES
INTERNATIONALBUSINESSMACINES
INTERNATIONALBUSINESSMACJINESCOPORATION
INTERNATIONALBUSINESSMAHINES
INTERNATIONALBUSINESSSMACHINES
INTERNATIONALBUSINSESSMACHINES
INTERNATIONALBUSINSSMACHINES
INTERNATIONALBUSISNESSMACHINES
INTERNATIONALBUSNESSMACHINES
INTERNATIONALBUSSINESSMACHINES
INTERNATIONLBUSINESSMACHINES
INTERNATONALBUSINESSMACHINES
INTERNTIONALBUSINESSMACHINES
INTRNATIONALBUSINESSMACHINES
ITERNATIONALBUSINESSMACHINES

LNTERNATIONALBUSINESSMACHINES

A considerable number of name variations were identified (n=74); the fact that they are all correct illustrates the potential power of approximate string searching.

In a second exercise, a larger scale test was performed on a sample of 1,000 cleaned names with a length of 15 characters. All Levenshtein distances between the names in the sample and all cleaned names with a length between 14 or 16 characters were calculated. 291 links were found with a Levenshtein distance of 1. Table 3 contains 100 matches with a Levenshtein distance of 1.

**Table 3: Example of 100 approximate string matches with Levenshtein distance of 1**

| SOURCE NAME | MATCHED NAME | SOURCE NAME | MATCHED NAME |
|---|---|---|---|
| AGINTERNATIONAL | AHINTERNATIONAL | APMANUFACTURING | CPMANUFACTURING |
| AGINTERNATIONAL | ALINTERNATIONAL | ALTEATECHNOLOGY | ALTHEATECHNOLOGY |
| ABAYATECHNOLOGY | AVAYATECHNOLOGY | APMANUFACTURING | AMMANUFACTURING |
| ABAYENGINEERING | ABALENGINEERING | ABKNUFSILPLATAR | ABKNUTSILPLATAR |
| ALINTERNATIONAL | AILINTERNATIONAL | ALPHATERAPEUTIC | ALPHATHERAPEUTIC |
| ALINTERNATIONAL | AHINTERNATIONAL | ABCOENGINEERING | AMCOENGINEERING |
| ABBALSTROMPOWER | ABBALSTOMPOWER | AGINTERNATIONAL | AZINTERNATIONAL |
| ALINTERNATIONAL | ALSINTERNATIONAL | AHINTERNATIONAL | ACINTERNATIONAL |
| ALINTERNATIONAL | ACINTERNATIONAL | ABCSCHOOLSUPPLY | ABCESCHOOLSUPPLY |
| APPLIEDMATRIALS | APPLIEDMATERIALS | ALIENTECHNOLOGY | ALLENTECHNOLOGY |
| APPLIEDMATRIALS | APPLIEDMATEIALS | APPLIEDGENETICS | APPLIEDGENERICS |
| ABBOTLABORATORY | ABBOTTLABORATORY | AMMANUFACTURING | APMANUFACTURING |
| APPLIEDMATERIAL | APPLIEDMATERIALS | ALIENTECHNOLOGY | ALIGNTECHNOLOGY |
| AGINTERNATIONAL | AGTINTERNATIONAL | AAMANUFACTURING | APMANUFACTURING |
| APPLIEDMATEIALS | APPLIEDMATRIALS | ABEAMTECHNOLOGY | BEAMTECHNOLOGY |
| APPLIEDMATEIALS | APPLIEDMATERIALS | ABEAMTECHNOLOGY | YBEAMTECHNOLOGY |
| ALINTERNATIONAL | AGINTERNATIONAL | AGINTERNATIONAL | ACINTERNATIONAL |
| ALINTERNATIONAL | CLINTERNATIONAL | AMMANUFACTURING | A1MANUFACTURING |
| AAMANUFACTURING | BAMANUFACTURING | ALPINEPLANTFOOD | ALPINEPLANTFOODS |
| AAMANUFACTURING | DAMANUFACTURING | AMMANUFACTURING | AAMANUFACTURING |
| AAMANUFACTURING | GAMANUFACTURING | AMMANUFACTURING | AEMANUFACTURING |
| AAMANUFACTURING | TAMANUFACTURING | APPLIEDGENERICS | APPLIEDGENETICS |
| AGINTERNATIONAL | AMINTERNATIONAL | 3DINTERNATIONAL | LDINTERNATIONAL |
| APTEKTECHNOLOGY | OPTEKTECHNOLOGY | AAMANUFACTURING | AMMANUFACTURING |
| ALINTERNATIONAL | TLINTERNATIONAL | AMINTERNATIONAL | PMINTERNATIONAL |
| ALINTERNATIONAL | AMINTERNATIONAL | 3DINTERNATIONAL | 3DMINTERNATIONAL |
| ALINTERNATIONAL | LLINTERNATIONAL | 3DINTERNATIONAL | 3EINTERNATIONAL |
| AGINTERNATIONAL | AGRINTERNATIONAL | 3DINTERNATIONAL | 3RINTERNATIONAL |
| ALINTERNATIONAL | CALINTERNATIONAL | 3DINTERNATIONAL | A3DINTERNATIONAL |
| ALINTERNATIONAL | AZINTERNATIONAL | AMINTERNATIONAL | PAMINTERNATIONAL |
| AAWPRODUCKTIONS | AAWPRODUKTIONS | 3DINTERNATIONAL | IDINTERNATIONAL |
| ALINTERNATIONAL | AOLINTERNATIONAL | 3DDEVELOPPEMENT | ADDEVELOPPEMENT |
| APPRICHJOHANNES | DAPPRICHJOHANNES | 3EINTERNATIONAL | 3AINTERNATIONAL |
| ALIGNTECHNOLOGY | ALIENTECHNOLOGY | 3EINTERNATIONAL | 3DINTERNATIONAL |
| ABALENGINEERING | ABAYENGINEERING | AMINTERNATIONAL | TAMINTERNATIONAL |
| ALINTERNATIONAL | PALINTERNATIONAL | 3EINTERNATIONAL | 3RINTERNATIONAL |
| AHINTERNATIONAL | ALINTERNATIONAL | 3EINTERNATIONAL | CEINTERNATIONAL |
| APPLIEDMAGNETIC | APPLIEDMAGNETICS | 3EINTERNATIONAL | GEINTERNATIONAL |
| AMMANUFACTURING | JMMANUFACTURING | 3EINTERNATIONAL | MEINTERNATIONAL |
| AMMANUFACTURING | LMMANUFACTURING | 3DINTERNATIONAL | FDINTERNATIONAL |
| AMMANUFACTURING | MMMANUFACTURING | 3AINTERNATIONAL | 3RINTERNATIONAL |
| AMMANUFACTURING | PMMANUFACTURING | ARANYKALSZMGTSZ | ARANYKALASZMGTSZ |
| AMINTERNATIONAL | ACINTERNATIONAL | AMINTERNATIONAL | ASMINTERNATIONAL |
| AHINTERNATIONAL | AGINTERNATIONAL | AMINTERNATIONAL | ATMINTERNATIONAL |
| ABKNUTSILPLATAR | ABKNUFSILPLATAR | AMINTERNATIONAL | AZINTERNATIONAL |
| APPIEDMATERIALS | APPLIEDMATERIALS | AMINTERNATIONAL | DMINTERNATIONAL |
| AHINTERNATIONAL | AMINTERNATIONAL | AMINTERNATIONAL | IMINTERNATIONAL |
| ALTEATECHNOLOGY | ALTELTECHNOLOGY | 3DINTERNATIONAL | 3AINTERNATIONAL |
| ALTEATECHNOLOGY | ALTEXTECHNOLOGY | 3AINTERNATIONAL | 3EINTERNATIONAL |
| APMANUFACTURING | FPMANUFACTURING | 3EINTERNATIONAL | SEINTERNATIONAL |

An inspection of Table 3 reveals that, in several cases, names cannot be assumed to refer to one and the same patentee automatically (e.g. "AG INTERNATIONAL", "AH INTERNATIONAL" and "AL INTERNATIONAL", or "APPLIED GENERICS" and "APPLIED GENETICS"). In other words, approximate string searching is very powerful in identifying potential matches but does not result in conclusive findings. The difference with the former example in Table 2 is the presence of proper names. Approximate string searching is conclusive when identifying spelling variations of common words, but far less conclusive in the case of proper names. Without additional validation efforts, the number of mismatches can be considerable. This problem will be more marked, the shorter the length of strings being assessed. This lack of accuracy precludes the adoption of approximate string searching in an automated manner. This issue might be addressed by using address information as discussed in section 6.3 – Introducing address information (in conjunction with name similarity)[13].

## 6.2 Automatic acronym generation

As mentioned in the introduction, the presence and use of acronyms and abbreviations result in an increase in name variations. To deal with acronyms, an automated method of generating acronyms for company names that consist of different parts might be considered. These generated acronyms can be matched with acronyms already present in the list of patentee names. When a match is found, one could consider harmonizing both names. This, however, requires the generated acronyms to be unambiguously related to an acronym already present. Unfortunately, this is rarely the case as the following experiment illustrates.

Acronyms have been generated automatically, beginning with a test set of names consisting of different parts (containing at least one space). The name after its legal form indication had been removed was used as the starting point. All non-letter (A-Z) and non-digit (0-9) characters were replaced with a space (" "), resulting in a string of words separated by spaces. An acronym was generated taking the first character of every word. These acronyms can be linked back to the cleaned names to see if automatically generated acronyms match acronyms already present in the name list.

Table 4 and Table 5 contain all names in the test set for which the automatic created acronym resulted in "IBM" and "ICC" respectively.

**Table 4: Names with automatically generated acronym "IBM"**

| AUTOMATICALLY GENERATED ACRONYM | ORIGINAL NAME |
|---|---|
| IBM | IBM BUSINESS MACHINES |
| IBM | IINTERNATIONAL BUSINESS MACHINES |
| IBM | INDUSTRIEANLAGEN BETRIEBSGESELLSCHAFT MBH |
| IBM | INDUSTRIEANLAGEN-BETRIEBSGESELLSCHAFT MBH |
| IBM | INDUSTRIEANLAGEN-BETRIEBTGESELLSCHAFT MBH |
| IBM | INERNATIONAL BUSINESS MACHINES |
| IBM | INFORMATION BUSINESS MACHINES |
| IBM | INTELLECTUAL BUSINESS MACHINES |
| IBM | INTENATIONAL BUSINESS MACHINES |
| IBM | INTERANATIONAL BUSINESS MACHINES |
| IBM | INTERANTIONAL BUSINESS MACHINES |
| IBM | INTERNAIONAL BUSINESS MACHINES |
| IBM | INTERNAITONAL BUSINESS MACHINES |
| IBM | INTERNAL BUSINESS MACHINE |
| IBM | INTERNATIAONAL BUSINESS MACHINES |
| IBM | INTERNATIIONAL BUSINESS MACHINES |
| IBM | INTERNATINAL BUSINESS MACHINES |
| IBM | INTERNATIOAL BUSINESS MACHINES |
| IBM | INTERNATIOANAL BUSINESS MACHINES |
| IBM | INTERNATIOANL BUSINESS MACHINES |
| IBM | INTERNATIOINAL BUSINESS MACHINES |

[13] For example, according to EPO applicant information, "APPLIED GENERICS" is always situated in Biggar, GB, while "APPLIED GENETICS" is always situated in Freeport, US, suggesting two different companies.

| | |
|---|---|
| IBM | INTERNATION BUSINESS MACHINES |
| IBM | INTERNATIONAAL BUSINESS MACHINES |
| IBM | INTERNATIONAL BOOK MARKETING |
| IBM | INTERNATIONAL BUINESS MACHINES |
| IBM | INTERNATIONAL BUISNESS MACHINES |
| IBM | INTERNATIONAL BUNSINESS MACHINES |
| IBM | INTERNATIONAL BUSIENSS MACHINES |
| IBM | INTERNATIONAL BUSINBESS MACHINES |
| IBM | INTERNATIONAL BUSINEES MACHINES |
| IBM | INTERNATIONAL BUSINEMSS MACHINES |
| IBM | INTERNATIONAL BUSINES MACHINES |
| IBM | INTERNATIONAL BUSINESS MACHIENS |
| IBM | INTERNATIONAL BUSINESS MACHINCES |
| IBM | INTERNATIONAL BUSINESS MACHINE |
| IBM | INTERNATIONAL BUSINESS MACHINES |
| IBM | INTERNATIONAL BUSINESS MACHINESC |
| IBM | INTERNATIONAL BUSINESS MACHINESS |
| IBM | INTERNATIONAL BUSINESS MACHNES |
| IBM | INTERNATIONAL BUSINESS MACHNIES |
| IBM | INTERNATIONAL BUSINESS MACHNINES |
| IBM | INTERNATIONAL BUSINESS MACINES |
| IBM | INTERNATIONAL BUSINESS MAHCINES |
| IBM | INTERNATIONAL BUSINESS MAHINES |
| IBM | INTERNATIONAL BUSINESS MCAHINES |
| IBM | INTERNATIONAL BUSINESSS MACHINES |
| IBM | INTERNATIONAL BUSINSESS MACHINES |
| IBM | INTERNATIONAL BUSINSS MACHINES |
| IBM | INTERNATIONAL BUSISNESS MACHINES |
| IBM | INTERNATIONAL BUSNESS MACHINES |
| IBM | INTERNATIONAL BUSSINESS MACHINES |
| IBM | INTERNATIONAL, BUSINESS MACHINES |
| IBM | INTERNATIONL BUSINESS MACHINES |
| IBM | INTERNATONAL BUSINESS MACHINES |
| IBM | INTERNTIONAL BUSINESS MACHINES |
| IBM | INTRNATIONAL BUSINESS MACHINES |
| IBM | ITERNATIONAL BUSINESS MACHINES |

**Table 5: Names with automatically generated acronym "ICC"**

| AUTOMATICALLY GENERATED ACRONYM | ORIGINAL NAME |
|---|---|
| ICC | I.C. COM |
| ICC | I.C. CONSULTANTS |
| ICC | IMPERIAL CHEMICAL COMPANY |
| ICC | INDIANA CARTON COMPANY |
| ICC | INDUSTRIAL CHEMICAL CONSULTANTS |
| ICC | INDUSTRIAL CONVEYOR COMPANY |
| ICC | INDUSTRIE CHIMICHE CAFFARO |
| ICC | INDUSTRIE CHIMICHE CAPPARO |
| ICC | INGRAM CACTUS COMPANY |
| ICC | INNOVATIVE CLEANING CONCEPTS |
| ICC | INNOVATIVE CULINARY CONCEPTS |
| ICC | INSPIRED & CREATED CONCEPTS |
| ICC | INTEGRATED CONTROL CONCEPTS |
| ICC | INTERMOUNTAIN CANOLA COMPANY |
| ICC | INTERNATIONAL CHEMICAL CONSULTANT |
| ICC | INTERNATIONAL CLAMP COMPANY |
| ICC | INTERNATIONAL CONNECTORS & CABLE |

Whilst the number of relevant matches is considerable in the case of "IBM", the same does not hold for the "ICC" example. Clearly, even the IBM example contains 'false' matches. In other words, both examples demonstrate that automated application of this approach will negatively affect accuracy. In order to avoid this negative impact, additional validation efforts are needed. Hence, while potentially useful, this method's lack of accuracy when applied in an automated way militates against its inclusion in this methodology.

## 6.3 Introducing address information (in conjunction with name similarity)

When engaging in name harmonizing efforts, it seems natural to consider the inclusion of address information of patentees such as country code, city name, zip/post code and street information. Address information can be used both for the additional identification of name variations (patentees with partly different names but identical addresses) and for the identification of potential mismatches (patentees having similar names but different addresses)[14].

Indeed, when both patentees share the same address, one might examine the possibility of harmonizing them. Given sufficient levels of name similarity (to avoid mismatches when different organizations share the same premises), there would appear to be a high probability that patentees are identical in these cases. In order to assess whether such an extension would be feasible, an analysis was performed to verify whether robust and unambiguous criteria could be outlined. Consequently, we examined a sample of 5,000 EPO applications. Names have been cleaned as described in section 4 - A content-driven name harmonization approach focusing on accuracy - and addresses have been cleaned in a very preliminary way (removal of all non-alphabetical characters). For those patentee names having the same address (country, city and street), the Levenshtein distance measure has been calculated and normalized for the varying lengths of names (absolute Levenshtein distance divided by the length of the longest names). The obtained matches have been verified in terms of correctness (is it reasonable, based on a quick verification of patentee information found, to assume that both patentees are one and the same?).

Table 6 contains the 25 cleaned patentee names with the same address with the closest relative Levenshtein distance out of the sample of 5,000 names. The absolute Levenshtein distance and the result of the validation is also included: "=" means that names are variants; "≈" means that names definitely have some relationship but not clear if it is the same legal entity; and "≠" means that matched names are significantly different (but still can point to the same legal entity; name can be significantly different because of name changes or mergers and acquisitions).

**Table 6: Differences in cleaned names of patentees with matched address**

| CLEANED NAME | CLEANED NAME WITH MATCHED ADRESS | ABS DIST | REL DIST | VAL |
|---|---|---|---|---|
| SCHNEIDERELECTRICINDUSTRYSAS | SCHNEIDERELECTRICINDUSTRY | 3 | 0,11 | = |
| MERRELLPHARMACEUTICALS | MERRELLDOWPHARMACEUTICALS | 3 | 0,12 | ≈ |
| MITSUBISHICHEMICAL | MITSUBISHIGASCHEMICAL | 3 | 0,14 | ≈ |
| THGOLDSCHMIDT | GOLDSCHMIDT | 2 | 0,15 | ≠ |
| COSMAINTERNATIONAL | MAGNAINTERNATIONAL | 4 | 0,22 | ≠ |
| KUMIAICHEMICALINDUSTRY | IHARACHEMICALINDUSTRY | 5 | 0,23 | ≠ |
| TAKEDACHEMICALINDUSTRY | WAKOPURECHEMICALINDUSTRY | 6 | 0,25 | ≠ |
| SUMITOMOMETALINDUSTRY | SUMITOMOELECTRICINDUSTRY | 6 | 0,25 | ≈ |
| ACCENTURELLP | ACCENTURE | 3 | 0,25 | = |
| MITSUBISHIDENKI | MITSUBISHIKASEI | 4 | 0,27 | ≈ |
| SHIBANAIAKIKO | SHIBANAIHIROKO | 4 | 0,29 | ≈ |
| RHONEPOULENCRORER | RHONEPOULENCSANTE | 5 | 0,29 | ≈ |
| GECMARCONI | THEMARCONI | 3 | 0,30 | ≈ |
| FORDGLOBALTECHNOLOGY | VISTEONGLOBALTECHNOLOGY | 7 | 0,30 | ≠ |
| BOEHRINGERINGELHEIM INTERNATIONAL | BOEHRINGERINGELHEIMVETMEDICA | 10 | 0,31 | ≈ |
| AUGWINKHAUS | FIRMAAUGWINKHAUS | 5 | 0,31 | = |
| SGSTHOMSONMICROELECTRONIC | STMICROELECTRONIC | 8 | 0,32 | ≈ |
| MITSUBISHIGASCHEMICAL | MITSUBISHIKASEI | 7 | 0,33 | ≈ |
| ASAHIKASEIKOGYO | ASAHIKASEI | 5 | 0,33 | ≈ |
| JOHNSONJOHNSONCLINICAL | ORTHOCLINICALDIAGNOSTICS | 11 | 0,33 | ≠ |

[14] In addition, similar addresses appearing jointly with different patentee names might trigger an assessment of ownership relationships. Such an approach can be beneficial to support legal entity harmonization efforts. However, as explained in section 2 - Patentee name harmonization and legal entity harmonization, this lies outside the scope of the methodology outlined in this paper, which is aimed at name harmonization.

| | | | | |
|---|---|---|---|---|
| DIAGNOSTICS | | | | |
| MITSUBISHIJUKOGYO | MITSUBISHIKASEI | 6 | 0,35 | ≈ |
| ALLERGANSALES | ALLERGAN | 5 | 0,38 | ≈ |
| CATERPILLAR | CATERPILLARTRACTOR | 7 | 0,39 | ≈ |
| KONINKLIJKEPHILIPSELECTRONIC | PHILIPSELECTRONIC | 11 | 0,39 | = |
| CENTRENATIONALDELARECHERCHE SCIENTIFIQUE | ETABLISSEMENTPUBLICDITCENTRE NATIONALDELARECHERCHE SCIENTIFIQUECNRS | 26 | 0,40 | = |

If correct matches were to coincide with distinctive values in terms of the distance measures, automated procedures could be envisaged. However, the results of this exercise do not suggest such a pattern. An inspection of the distance values and validation reveals that defining an unambiguous criterion value is not so straightforward. Moreover, one observes immediately that several cases require a more in-depth analysis in order to define whether or not both patentees are similar. Such an assessment immediately extends beyond name harmonizing per se and is best categorized as an exercise in legal entity harmonization (see section 2 – Patentee name harmonization and legal entity harmonization). As this approach is clearly beyond the scope of the objectives envisaged in this contribution, it has not been included in the final methodology. At the same time, it goes without saying that enriching approximate string searching with address information seems highly promising when striving for legal entity harmonization and might also contribute to automated name harmonizing. Both extensions do imply however considerable validation efforts.

# 7 CONCLUSION

In this contribution, we have developed a comprehensive approach oriented towards name harmonizing. Emphasis has been placed on maximizing the accuracy of procedures that can be implemented automatically, i.e. without additional - and time-consuming - validation efforts that require secondary information sources. Name variations are not combined if there is any doubt that the names relate to different legal entities.

This has resulted in a transparent method whose outcome has been a reduced set of harmonized names. At this stage, the total number of patentee names has been decreased by 17%. 13.5% of the harmonized names are matched with more than one original patentee name, matching 2.6 names on average. When harmonizing takes place, the number of patents allocated to the same entity increases on average by 7.2 patents in absolute terms (33.1% in relative terms), signaling a considerable impact.

A detailed validation exercise was conducted for 35 harmonized names. Findings revealed accuracy levels of 100% percent and a level of completeness of 99,62%. More details about this validation can be found in Appendix 7.

EUROSTAT and its partners deliberately opted for a transparent method so that all interested parties will be able to build further on the results obtained. In the belief that the procedures described in this methodology can be further enriched and refined – and this also applies to legal entity normalization - we would encourage activities in this direction.

Improving the accuracy levels for the methodology as a whole is feasible by introducing expert assessments in a systematic manner. Given the volume of names involved, such an effort is beyond the current resources of EUROSTAT and its partners who developed this methodology (INCENTIM/SOOS, K.U.Leuven, and SOGETI). At the same time, numerous researchers and analysts are currently working on name harmonizing efforts with specific samples (e.g. technological fields, countries/regions, and sectors). For researchers engaged in such efforts, building on this methodology might be helpful; equally, the insights obtained by researchers and analysts might be beneficial for further refinement of the current methodology. In other words, by sharing the methodology developed among the different communities involved in patentee analysis, further improvements could be envisaged. Consequently, EUROSTAT and its partners decided to put the complete methodology into the public domain.

In Appendix 1 and Appendix 2, the full set of procedures is made available, making allowances for implementation, verification, and the development of appropriate extensions. Furthermore, given its continuous involvement in the PATSTAT Taskforce activities, EUROSTAT, in collaboration with the researchers at K.U.Leuven who developed this methodology, is committed to making freely available all future improvements in this methodology, including those obtained from other researchers and analysts[15].

[15] A toolbox is available from the authors with the automated procedure and a framework for future extensions.

# 8 REFERENCES


- DERWENT WORLD PATENTS INDEX Patentee Codes, Revised Edition 8, 2002, Thomson Scientific, United Kingdom, ISBN 0 901157 38 4 (www.thomsonscientific.com/media/scpdf/patenteecodes.pdf)
- Griliches, Z. (1990). "Patent statistics as economic indicators: A survey." Journal of Economic Literature, 28, 1661-1707
- Johnson, D. K. N. (2002). "The OECD Technology Concordance (OTC): patents by industry of manufacture and sector of use." STI Working Papers 2002/5
- Schmoch U., Laville F., Patel P., Frietsch R. (2003). "Linking Technology Areas to Industrial Sectors" Final Report to the European Commision, DG Research
- Verspagen B., van Moergastel T., Slabbers M. (1994). "MERIT concordance table: IPC – ISIC (rev. 2)" MERIT Research Memorandum 2/94-004
- Wu S. and Manber U. (1992). "AGREP -- A fast approximate pattern-matching tool" Proc. Winter 1992 USENIX Technical Conference, 153-162

# APPENDIX 1: STEP-BY-STEP METHODOLOGY AND APPLICATION USING EPO AND USPTO PATENTEE NAMES

This appendix describes in detail all steps in the name-cleaning and harmonization procedures.

A description is given of the objective for each generic step, followed by an analysis of the patentee name set to find the best way to achieve the objectives of the step at hand, bearing a maximum level of accuracy in mind.

Given the results of the data specific analysis, an implementation procedure is proposed to automate the cleaning step.

Finally, the results of the implementation of the cleaning step on the dataset with patentee names are given, along with the impact on the number of unique names in the set.

The dataset used to develop and test the name-cleaning and harmonization procedures is a combination of EPO applicant names and USPTO assignee names.

For EPO data, all 270,635 applicant names are included from all 1,600,812 EPO patent applications published between 1978 and 2004 (based on the EPO ESPACE ACCESS product). All names were converted to uppercase.

For USPTO data, all 223,665 assignee names are included from all 1,614,224 USPTO granted patents published between 1991 and 2003 (based on the USPTO Grant Red Book product). All names were converted to uppercase.

Combining those two datasets resulted in a name list of 443,722 unique patentee names (after conversion to uppercase and removal of leading and trailing spaces). Of these names, 50,578 appear both in EPO and USPTO; 173,087 only appear in USPTO, while 220,057 only appear in EPO.

# 1 DATA PRE-PROCESSING

## 1.1 Character cleaning

### 1.1.1 Remove HTML codes

#### Description

HTML codes such as "<BR>" (single-line break) are only relevant to the formatting of names. These codes can be removed as they are of no significance in a name.

#### Analysis

HTML codes are identified by querying the data for the following pattern: "%<%>%".

Not all query results have to be HTML codes but the query result can be used to identify all occurring HTML codes.

Only the "<BR>" HTML tag is found in the data. This tag defines a single line break in HTML and can be replaced with a space in a name.

#### Implementation

All occurrences of "<BR>" are replaced with a space " " by executing an update query on the data.

As replacement with a space can lead to leading or trailing spaces, names have to be checked for and trimmed of leading and trailing spaces after the removal of HTML codes.

#### Results

"<BR>" has been replaced with a space in 8,874 names.

No other HTML codes are present in the names.

#### Impact

From 443,722 unique names to 442,795 unique names, a reduction of 927 names (0.2%).

### 1.1.2 Replace SGML coded characters

#### Description

SGML coded characters such as "&" or "&OACUTE;" should be replaced with their normal ASCII/ANSI equivalent, whenever possible.

#### Analysis

SGML coded characters are identified by querying the data for the following pattern: "%&%;%".

Not all query results are real SGML coded characters. For example, "HITACHI ENGINEERING & SERVICES CO; LTD." matches the pattern but no SGML coded character is involved. However, the query result can be used to identify all occurring SGML coded characters.

Table 7 contains the SGML coded characters that were found in the names.

**Table 7: SGML codes and their ASCII/ANSI equivalent**

| SGML CODE | REPLACEMENT CHARACTER |
|---|---|
| & | & |
| &OACUTE; | Ó |
| &SECT; | § |
| &UACUTE; | Ú |
| ⋆ | replace with space |
| &BULL; | . |
| &EXCL; | ! |

#### Implementation

All occurrences of SGML coded characters are replaced with their respective ASCII/ANSI equivalent, as defined in Table 7, by executing several update queries on the data.

The order of the replacement is important, especially in the case of the "&" SGML code. Every SGML code starts with an ampersand but, sometimes, this ampersand, as part of an SGML code, is also represented by the SGML code "&". For example, the following code might appear: "&EXCL;". This is, in fact, the SGML code for an exclamation mark ("&EXCL;") but with the first ampersand also coded as an SGML character. These kinds of codes are correctly converted if, first of all, the "&AMP" code is replaced with "&", resulting in code "&EXCL;" that can be replaced with "!". The "&" code must always be replaced first, before other codes.

As replacement with a space can result in leading or trailing spaces, names have to be checked for and trimmed of leading and trailing spaces after replacement of SGML coded characters.

#### Results

SGML coded characters have been replaced by their ASCII/ANSI equivalent in 12,430 names.

No other SGML coded characters are present in the names.

#### Impact

From 442,795 unique names to 440,237 unique names, an additional reduction of 2,558 names, or a total reduction of 3,485 names (0.8%).

### 1.1.3 Replace propriety coded characters

#### Description

In addition to SGML character coding, other proprietary character coding can be used by data suppliers to code special characters.

For USPTO data, codes like "{UMLAUT OVER (A)}" and "{DOT OVER (E)}" can be found. These coded characters should be replaced with their normal ASCII/ANSI equivalents whenever possible.

#### Analysis

Proprietary coded characters are identified by querying the data for the following pattern: "%{%}%"; "%[%]%" and "%(%)%".

Not all query results have to be proprietary coded characters but the query result can be used to identify all occurring proprietary coded characters.

Table 8 contains the proprietary coded characters that were found in the names.

**Table 8: Proprietary character codes and their ASCII/ANSI equivalent**

| PROPRIETARY CODED CHARACTER | REPLACEMENT CHARACTER |
|---|---|
| "{UMLAUT OVER (A)}" | "Ä" |
| "{UMLAUT OVER (E)}" | "Ë" |
| "{UMLAUT OVER (O)}" | "Ö" |
| "{UMLAUT OVER (U)}" | "Ü" |
| "{UMLAUT OVER (N)}" | "N" |
| "{UMLAUT OVER (R)}" | "R" |
| "{UMLAUT OVER (Z)}" | "Z" |
| "{ACUTE OVER (A)}" | "Á" |
| "{ACUTE OVER (E)}" | "É" |
| "{ACUTE OVER (T)}" | "T" |
| "{ACUTE OVER (V)}" | "V" |
| "{GRAVE OVER (B)}" | "B" |
| "{GRAVE OVER (R)}" | "R" |
| "{OVERSCORE (A)}" | "A" |
| "{OVERSCORE (D)}" | "D" |
| "{OVERSCORE (E)}" | "E" |
| "{OVERSCORE (O)}" | "O" |
| "{OVERSCORE (U)}" | "U" |
| "{DOT OVER (A)}" | "A" |
| "{DOT OVER (E)}" | "E" |
| "{DOT OVER (U)}" | "U" |
| "{HAECK OVER (C)}" | "C" |
| "{HAECK OVER (S)}" | "S" |

#### Implementation

All occurrences of proprietary coded characters are replaced with their respective ASCII/ANSI equivalent, as defined in Table 8, by executing several update queries on the data.

#### Results

Proprietary character codes have been replaced with their ASCII/ANSI equivalent in 62 names.

The possibility cannot be ruled out that other proprietary character codes are still present in the names.

#### Impact

From 440,237 unique names to 440,206 unique names, an additional reduction of 31 names, or a total reduction of 3,516 names (0.8%).

### 1.1.4 Replace accented characters

#### Description

Some foreign languages such as French and German use accented characters like "é" or "ü". However, these accented characters are not always used in the spelling of names and, therefore, should be replaced with their unaccented equivalents.

For some characters in some languages, replacement of accented characters with unaccented equivalents might not be straightforward because of alternative spelling. For example, characters in German (and in other languages such as Hungarian) with a diacritic mark ('umlaut': "ä", "ö", "ü") have an alternative spelling with an additional "e"; therefore, the real equivalent of "wärme" is "waerme", the real equivalent of "förderung" is "foerderung" and the real equivalent of "für" is "fuer". However, in practice, the simplified spelling can also be found, so both "für", "fuer" and "fur" can appear. This raises the question whether the German "ä", "ö" and "ü" should be replaced by "a", "o" and "u" or "ae", "oe" and "ue" respectively.

As the simplified spelling (without the additional "e") already appears in the original names, in addition to the accented spelling and the real equivalent with the additional "e", all accented characters are replaced with their simple underlying equivalent without an umlaut and without an additional "e". This means that, as a result of this cleaning step, "für" will be harmonized with "fur", but "fuer" will not be harmonized with "fur" or "für".

A separate step – Umlaut harmonization - will be used later on to harmonize language-specific spelling variations of accented characters.

#### Analysis

All kinds of accented characters can appear in different languages. An exhaustive list of all possible accented characters in all languages has not been used. Instead, this step focuses on those characters that can be represented as a single character, as defined by the standard ASCII/ANSI character code page. All other possible accented characters cannot always be represented correctly by all software, and are mostly coded (see above, SGML and proprietary coded characters).

Table 9 contains the accented characters that can be found in the ASCII/ANSI character code page and their unaccented variant (only uppercase characters are taken into account, as the dataset is supposed only to contain uppercase characters).

**Table 9: Accented characters and their unaccented equivalent**

| CODE | CHARACTER | UNACCENTED EQUIVALENT |
|---|---|---|
| 192 | "À" | "A" |
| 193 | "Á" | "A" |
| 194 | "Â" | "A" |
| 195 | "Ã" | "A" |
| 196 | "Ä" | "A" |
| 197 | "Å" | "A" |
| 198 | "Æ" | "AE" |
| 199 | "Ç" | "C" |
| 200 | "È" | "E" |
| 201 | "É" | "E" |
| 202 | "Ê" | "E" |
| 203 | "Ë" | "E" |
| 204 | "Ì" | "I" |
| 205 | "Í" | "I" |
| 206 | "Î" | "I" |
| 207 | "Ï" | "I" |
| 209 | "Ñ" | "N" |
| 210 | "Ò" | "O" |
| 211 | "Ó" | "O" |
| 212 | "Ô" | "O" |
| 213 | "Õ" | "O" |
| 214 | "Ö" | "O" |
| 217 | "Ù" | "U" |
| 218 | "Ú" | "U" |
| 219 | "Û" | "U" |

| | | |
|---|---|---|
| 220 | "Ü" | "U" |
| 221 | "Ý" | "Y" |
| 159 | "Ÿ" | "Y" |

#### Implementation

All occurrences of accented characters are replaced with their respective unaccented character equivalent, as defined in Table 9, by executing several update queries on the data.

#### Results

Accented characters have been replaced with their unaccented equivalent in 19,934 names.

There is no guarantee that single accented characters will be completely eliminated from the names, as an exhaustive list of all possible accented characters in all languages has not been used.

However, no other accented characters that can be represented as a single character, as defined by the standard ASCII/ANSI character code page, are present in the names.

#### Impact

From 440,206 unique names to 438,366 unique names, an additional reduction of 1,840 names, or a total reduction of 5,356 names (1.2%).

### 1.1.5 Check for special characters

#### Description

After replacement of SGML coded characters, proprietary coded characters and accented characters, no special characters should remain. 'Special' refers to a character that is not expected in a name because it is not a letter, a digit, or a regular punctuation character.

#### Analysis

Special characters are identified by querying the data for characters that are not part of the following set of letters, digits and punctuation characters: A-Z; 0-9; "-"; "+"; "'"; """; "#"; "*"; "@"; "!"; "?"; "/"; "&"; "("; ")"; ":"; ";"; ","; "."; " ".

63 names were found to contain special characters but none of them are problematic for the harmonization (in any case, non-alphanumerical characters and spaces are removed in a further step).

## 1.2 Punctuation cleaning (pre-parsing)

### 1.2.1 Replace double spaces

#### Description

Double spaces should be replaced with a single space.

#### Analysis

Double spaces are identified by querying the data for names having the pattern "%  %".

1,781 names were found to contain double spaces.

#### Implementation

All occurrences of double spaces are replaced with a single space by executing an update query on the data.

#### Result

Double spaces have been replaced in 1,781 names.

No other double spaces are present in the names.

#### Impact

From 438,366 unique names to 438,074 unique names, an additional reduction of 292 names, or a total reduction of 5,648 names (1.3%).

### 1.2.2 Remove double quotation mark irregularities

#### Description

Names beginning or ending with a double quotation mark should not contain a space after the beginning quotation mark or before the ending quotation mark respectively.

#### Analysis

Names that have a space after the beginning or before the ending double quotation mark are identified by querying the data for the following pattern: """ %" or "% """

29 names were identified that have spaces after the beginning or before the ending of the double quotation mark.

#### Implementation

All spaces after the beginning or before the ending of the double quotation mark are removed by executing several update queries on the data.

#### Results

Spaces after the beginning and before the ending double quotation marks have been removed in 29 names.

#### Impact

From 438,074 unique names to 438,069 unique names, an additional reduction of 5 names, or a total reduction of 5,653 names (1.3%).

### 1.2.3 Remove double quotation marks at the beginning and the end of a name

#### Description

Names that have a double quotation mark at the beginning and the end, and that do not contain any other double quotation mark, should have quotation marks removed.

#### Analysis

Names that have a double quotation mark and do not contain any other double quotation mark are identified by querying the data for the following pattern: “””%””” and not “””%””%”””.

52 names were identified that start and end with a double quotation mark and do not contain any other double quotation mark.

#### Implementation

All beginning and ending double quotation marks of names that start and end with a double quotation mark and do not have any additional double quotation marks were removed by executing an update query on the data.

Since the removal of double quotation marks at the beginning and the end of a name can lead to leading spaces, names have to be checked for and trimmed of leading spaces after removal of double quotation marks.

#### Results

Beginning and ending double quotation marks have been removed in 52 names.

#### Impact

From 438,069 unique names to 438,061 unique names, an additional reduction of 8 names, or a total reduction of 5,661 names (1.3%).

### 1.2.4 Remove non-alphanumerical characters at the beginning of a name

#### Description

A name is expected to begin with a letter, a digit, or some relevant character but not with a character such as ".” or ",”.

Non-alphanumerical characters at the beginning of a name that are not relevant should be removed.

#### Analysis

Names that begin with an irrelevant non-alphanumerical character are identified by querying the data for names where the first character does not belong to the following set of letters, digits and other relevant characters: A-Z; 0-9; "”"; "@"; "("; "'"; "#"; "!"; "*"; "/".

23 names were found to contain an irregular first character although only 18 actually begin with an irrelevant character.

The following characters were identified for removal if they appear at the beginning of a name: ".”; "-"; "?"; ":"; "_".

#### Implementation

All occurrences of ".”; "-"; "?"; "!"; "*"; ":"; "_" are removed from the beginning of a name by executing an update query on the data.

As the removal of irrelevant characters at the beginning of a name can lead to leading spaces, names have to be checked for and trimmed of leading spaces after the removal of irrelevant characters at the beginning of a name.

The removal of characters at the beginning of a name can also lead to a new irregular beginning of a name, so this step has to be executed several times until no further irregularities are found.

#### Result

Irrelevant non-alphanumerical characters at the beginning of a name have been removed in 18 names.

#### Impact

From 438,061 unique names to 438,052 unique names, an additional reduction of 9 names, or a total reduction of 5,670 names (1.3%).

### 1.2.5 Remove non-alphanumerical characters at the end of a name

#### Description

A name is expected to end with a letter, a digit or some relevant character, but not with a character like ":” or ";”.

Non-alphanumerical characters at the end of a name that are not relevant should be removed.

#### Analysis

Names that end with an irrelevant non-alphanumerical character are identified by querying the data for names where the last character does not belong to the following set of letters, digits and other relevant characters: A-Z; 0-9; ".”; "'"; "”"; ")".

1,528 names were found containing an irregular end.

668 end with " DITE:”, " DITE,” or " DITE :”. Normally, this should be followed by an acronym or nickname. Since this is not the case here, the 668 occurrences of "DITE:”, " DITE,” and "DITE :” can be removed.

The following characters were identified for removal if they appear at the end of a name: ",";
";"; ":"; "-".

#### Implementation

Firstly, all occurrences of " DITE:", " DITE," and "DITE :" at the end of names are removed by executing an update query on the data.

Next, all occurrences of ",", ";"; ":"; "-" are removed at the end of a name by executing an update query on the data.

As the removal of irrelevant characters at the end of a name can lead to trailing spaces, names have to be checked for and trimmed of trailing spaces after removal of irrelevant characters at the end of a name.

The removal of characters at the end of a name can also lead to a new irregular name ending, so this step has to be executed several times until no further irregularities are found.

#### Result

Irrelevant non-alphanumerical characters at the end of a name have been removed in 1,498 names.

#### Impact

From 438,052 unique names to 437,689 unique names, an additional reduction of 363 names, or a total reduction of 6,033 names (1.4%).

### 1.2.6 Replace comma irregularities

#### Description

A comma should be followed by a space and not be preceded by a space. A comma not followed by a space or preceded by a space means some irregularity in most cases.

#### Analysis

Firstly, comma irregularities based on commas not followed by a space are identified by querying the data for names having the pattern "%,[! ]%".

624 names were identified having a comma not followed by a space.

A pattern and case-based approach is used to clean irregularities instead of blindly adding a space after every comma not having a space. A fully automated approach is dangerous because a comma might be a decimal, a thousand separator, or can appear as an abbreviation indicator instead of a dot.

Table 10 contains most occurring patterns containing a comma not followed by a space found in the names.

**Table 10: Patterns with comma not followed by space**

| PATTERN | REPLACE WITH |
|---|---|
| "% CO.,LTD.%" | " CO., LTD." |
| "% CO.,LTD%" | " CO., LTD" |
| "% CO,. LTD.%" | " CO., LTD." |
| "% CO.,INC.%" | " CO., INC." |
| "%,LTD.%" | ", LTD." |
| "%,LTD" | ", LTD" |
| "%,INC.%" | ", INC." |
| "%,INC" | ", INC" |
| "%,LLC.%" | ", LLC." |
| "%,LLC" | ", LLC" |
| "%,L.L.C.%" | ", L.L.C." |
| "%,S.A.R.L.%" | ", S.A.R.L." |
| "%,S.A.%" | ", S.A." |
| "% CO,LTD" | " CO, LTD" |
| "% CO,KG.%" | " CO, KG." |
| "% CO.,KG" | " CO., KG" |

| | |
|---|---|
| "%,GMBH.%" | ", GMBH." |
| "%,GMBH" | ", GMBH" |
| "%,PLC" | ", PLC" |
| "%,S.R.L.%" | ", S.R.L." |

These patterns cover 396 cases. The other 228 cases are very diverse and difficult to capture in patterns and can be left unchanged because of the low numbers.

Next, comma irregularities based on commas preceded by a space are identified by querying the data for names having the pattern "% ,%".

89 names were identified as having a comma preceded by a space.

Table 11 contains most occurring patterns containing a comma preceded by a space found in the names.

**Table 11: Patterns with comma preceded by space**

| PATTERN | REPLACE WITH |
|---|---|
| "% , INC.%" | ", INC." |
| "% , LTD.%" | ", LTD." |
| "% , L.L.C.%" | ", L.L.C." |
| "% , LLC" | ", LLC" |
| "% , S.P.A.%" | ", S.P.A." |
| "% , S.A.%" | ", S.A." |

These patterns cover 49 cases. The other 40 cases can be corrected by replacing " ," with ",".

#### Implementation

Firstly, all occurrences of commas not followed by a space are replaced, as defined in Table 10, by executing several update queries on the data using the patterns found in the analysis.

Next, all occurrences of commas preceded by a space are replaced, as defined in Table 11, by executing several update queries on the data using the patterns found in the analysis.

Finally, all occurrences " ," are replaced with "," by executing an update query on the data.

#### Result

Commas not followed by a space have been replaced in 396 names.

Commas not preceded by a space have been replaced in 89 names.

Commas not followed by a space are still present in the names because the limited number of cases left are hard to clean automatically and hardly affect the cleaning and harmonizing steps that follow, and because commas can act as decimal or thousand separators.

#### Impact

From 437,689 unique names to 437,388 unique names, an additional reduction of 301 names, or a total reduction of 6,334 names (1.4%).

### 1.2.7 Replace period irregularities

#### Description

A period should be preceded by a letter or digit, and not another non-alphanumerical character or space. A period not preceded by a letter or digit means some irregularity in most cases and will be removed.

#### Analysis

Period irregularities based on periods not preceded by a letter or digit are identified by querying the data for names having the pattern "%[!A-Z0-9].%".

176 names were identified as having a period not preceded by a letter or digit.

Table 12 contains most occurring patterns containing a period not preceded by a letter or digit found in the names.

**Table 12: Patterns with period not preceded by a letter or digit**

| PATTERN | REPLACE WITH |
|---|---|
| "%, INC,." | ", INC." |
| "% CORP,." | " CORP." |
| "% CO,." | " CO." |
| "% COMPANY,. LIMITED" | " COMPANY, LIMITED" |
| % INC..% | " INC." |
| % S.A..% | " S.A." |
| % PTY. .LIMITED% | " PTY. LIMITED" |
| % CO.. INC.% | " CO. INC." |
| %, INC.. | ", INC." |
| % CO.. LTD. | " CO. LTD." |
| % A/.S% | " A/S" |
| % N..V | " N.V" |
| % LTD.. | " LTD." |
| % CO., LTD,. | " CO., LTD." |
| % CO., LTD.. | " CO., LTD." |
| % CO.., LTD. | " CO., LTD." |
| % P.L.C.. | " P.L.C." |

These patterns cover 55 cases. Most of the remaining cases are periods preceded by a space and can be corrected by replacing " ." with ".".

### Implementation

Firstly, all occurrences of periods not preceded by a letter or digit are replaced, as defined in Table 12, by executing several update queries on the data using the patterns found in the analysis.

Next, all occurrences of " ." are replaced with "." by executing an update query on the data.

### Result

Periods not preceded by a letter or digit have been replaced in 99 names.

Periods not preceded by a letter or digit are still present in the names because the limited number of cases left are hard to clean automatically and hardly affect the cleaning and harmonizing steps that follow, and because a series of periods "…" appear in the names to indicate that the name is abbreviated.

### Impact

From 437,388 unique names to 437,336 unique names, an additional reduction of 52 names, or a total reduction of 6,386 names (1.4%).

# 2 NAME CLEANING

## 2.1 Legal form indication treatment

### Description

A lot of organization names contain some kind of legal form indication (e.g. "INC.", "LIMITED", "LTD."). These legal form indications cause considerable name variation because of abbreviations, spelling variations, and legal form variations of names.

The name of a company can mostly be separated from the legal form without changing the real company name, although there are some exceptions of legal form that really are part of the name (see below). Moving and harmonizing legal forms to a separate field can greatly reduce the number of name variations.

The idea is to end up with the real name where non-relevant legal form indications are removed. It is not the intention to mutilate the organization name; the name still has to be complete and comprehensible. Whenever the legal form is part of the name, the legal form will not be removed.

For example, "S.A.B.C.A." or "SABCA" stands for "Société Anonyme Belge de Constructions Aéronautiques". "Société Anonyme" or "SA" is a legal form indication, but removing it from the name would leave "BCA" or "Belge de Constructions Aéronautiques", making the name hard to recognize.

This also means that if there is any doubt that part of the name is a legal form indication, the name should be left unchanged (some parts of a name can accidentally coincide with variations or abbreviations of a legal form).

The legal form indications are not completely deleted. They are removed from the name but, at the same time, the harmonized legal form is transferred to a different field. This gives the end user the opportunity to decide on whether two names that are identical except for the legal form should be considered the same entity.

For example, "IBM AG", "IBM INCORPORATED" and "IBM INC" will all be harmonized to "IBM" but, in a separate field, the first name will still be labeled as "AG" while the other two names will be labeled as "INCORPORATED", leaving the choice to the user to query on the harmonized name only (combining all three names) or to query on the combination of the harmonized name and the harmonized legal form (again splitting up the result between "IBM AG", on the one hand, and "IBM INCORPORATED", including "IBM INCORPORATED" and "IBM INC", on the other).

### Analysis

An official list of legal forms and their official abbreviations of all countries applying for patents can be a starting point for the identification of legal forms, but it is not very useful due to all kind of variations appearing in the patentee names.

An alternative approach is to index the last word of all organization names and check the top occurring words. All words that are not common English words are potentially legal form indications and can be checked in detail to see if they can be removed.

Table 13 contains the top 50 occurring last words after cleanup, along with the number of names containing the word as a last word, the cumulative number of names for this word and all higher ranked words, and the percentage of the cumulative number of names compared to the total number of names (443,722). Last words are identified on the basis of the last occurrence of a space in a name; then all non-(A-Z) and non-(0-9) characters are removed resulting in a cleaned version of the last word. Appendix 3 contains the list of the top 200 occurring last words.

**Table 13: Top 50 occurring last words**

| LAST WORD (CLEANED) | NBR OF NAMES | CUM | % | LAST WORD (CLEANED) | NBR OF NAMES | CUM | % |
|---|---|---|---|---|---|---|---|

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | INC | 74949 | 74949 | 17 | 26 | C | 1693 | 261265 | 59 |
| 2 | LTD | 35069 | 1001 | 25 | 27 | M | 1576 | 262841 | 59 |
| 3 | LIMITED | 20459 | 130477 | 29 | 28 | L | 1545 | 264386 | 60 |
| 4 | GMBH | 17490 | 147967 | 33 | 29 | OY | 1536 | 265922 | 60 |
| 5 | CORPORATION | 17348 | 165315 | 37 | 30 | E | 1446 | 267368 | 60 |
| 6 | SA | 9046 | 174361 | 39 | 31 | NV | 1376 | 268744 | 61 |
| 7 | KG | 7021 | 181382 | 41 | 32 | R | 1334 | 270078 | 61 |
| 8 | LLC | 6974 | 188356 | 42 | 33 | AKTIENGESELLSCHAFT | 1174 | 271252 | 61 |
| 9 | AG | 6786 | 195142 | 44 | 34 | W | 1142 | 272394 | 61 |
| 10 | SPA | 5967 | 201109 | 45 | 35 | D | 1120 | 273514 | 62 |
| 11 | CO | 5875 | 206984 | 47 | 36 | PETER | 1120 | 274634 | 62 |
| 12 | COMPANY | 5806 | 212790 | 48 | 37 | JOHN | 1118 | 275752 | 62 |
| 13 | SRL | 5501 | 218291 | 49 | 38 | PLC | 1077 | 276829 | 62 |
| 14 | | 5398 | 223689 | 50 | 39 | SARL | 1045 | 277874 | 63 |
| 15 | CORP | 5370 | 229059 | 52 | 40 | H | 1013 | 278887 | 63 |
| 16 | BV | 5165 | 234224 | 53 | 41 | MICHAEL | 988 | 279875 | 63 |
| 17 | AB | 4979 | 239203 | 54 | 42 | DIPLING | 957 | 280832 | 63 |
| 18 | INCORPORATED | 3671 | 242874 | 55 | 43 | S | 955 | 281787 | 64 |
| 19 | AS | 3168 | 246042 | 55 | 44 | G | 936 | 282723 | 64 |
| 20 | MBH | 2664 | 248706 | 56 | 45 | LP | 865 | 283588 | 64 |
| 21 | A | 2447 | 251153 | 57 | 46 | INTERNATIONAL | 865 | 284453 | 64 |
| 22 | DR | 2321 | 253474 | 57 | 47 | ROBERT | 814 | 285267 | 64 |
| 23 | KAISHA | 2208 | 255682 | 58 | 48 | P | 798 | 286065 | 64 |
| 24 | J | 2192 | 257874 | 58 | 49 | UNIVERSITY | 787 | 286852 | 65 |
| 25 | ANONYME | 1698 | 259572 | 58 | 50 | F | 778 | 287630 | 65 |

The top occurring word "INC" appears as last word in 74,949 names, the 50th occurring word "F" appears as the last word in 778 works. Together, these 50 words comprise 65% of all last words of all names.

The missing word in place 14 means that there are 5,398 words containing only one word (meaning containing no space). These names are supposed not to contain any legal form indication (because nothing would be left of the name), and the last words (the only word) of these names were not included in the last word index.

In the list, words can be found that clearly are legal form indications ("INC", "LTD", etc.), and there are those that clearly are not legal form indications ("PETER", "JOHN").

Some of the words are typically found in many company names but are not really legal form indications ("CORPORATION", "COMPANY").

All words having more than 1,000 occurrences are examined in detail, resulting in a list of 40 words, ranging from "INC" to "H" and covering 63% of all last words of all names.

For every word to be examined, the one but last and the second but last words were also cleaned and indexed to examine the combination of the three last words, as some legal form indications consist of more than one word or can be combined (e.g., the word "C" might not seem to be a legal form until one notices that the last but one word is "L" and the second but last word is "P", resulting in "PLC" or Public Limited Company).

Table 14 contains all last words identified as (being part of) legal form indications after examination of the last three word index.

**Table 14: Last words identified as legal form indications**

| LAST WORD (CLEANED) | LEGAL FORM |
|---|---|
| INC | Incorporated |
| LTD | Limited |
| LIMITED | Limited |
| GMBH | Gesellschaft mit beschränkter Haftung |
| SA | Société Anonyme, Sociedad Anónima, … |
| KG | Kommanditgesellschaft |
| LLC | Limited Liability Company |
| AG | Aktiengesellschaft |
| SPA | Società Per Azioni |
| SRL | Società a Responsabilità Limitata |
| BV | Besloten vennootschap |

| | |
|---|---|
| INCORPORATED | Incorporated |
| AS | Aktieselskab, Akciová Společnost |
| MBH | Gesellschaft mit beschränkter Haftung |
| A | Société Anonyme, Società Per Azioni, … |
| KAISHA | Kabushiki Kaisha |
| ANONYME | Société Anonyme |
| C | Pulic Limited Company, Limited Liability Company |
| L | Società a Responsabilità Limitata |
| OY | Osakeyhtiö |
| NV | Naamloze Vennootschap |
| AKTIENGESELLSCHAFT | Aktiengesellschaft |
| PLC | Public Limited Company |
| SARL | Société à responsabilité limitée |
| H | Gesellschaft mit beschränkter Haftung |

Out of the 40 words having more than 1,000 occurrences, 25 are identified as legal form indications or part of legal form indications. This does not mean that all those last words always indicate a legal form. Especially short words can be an abbreviation for any other word or can be an abbreviation of first name and middle names of private individuals. The context of the last word within the name has to be taken into consideration before blindly removing legal form indications. For example, "INCORPORATED" can always be removed but "A" will only be removed if it is preceded by "S " or "S. ", or "S P " or "S. P.".

So, for each of the 25 last words that potentially indicate a legal form, a thorough examination is needed to identify all spelling variations appearing at the end of names and to verify that a particular spelling variation really indicates a legal form and is not merely a coincidence.

For every identified last word, all names containing that word as the last word are scanned thoroughly to identify all spelling variations that actually indicate a legal form. Based on these examinations, search and replace rules could be constructed to remove and harmonize legal form indications.

For example, if a name ends with ", A.G.", this part of the name can safely by removed from the end of the name and another field can be updated to "AG" to indicate the harmonized legal form.

In total, 1,060 spelling variations of legal forms were identified at the end of names that can be safely removed.

Table 15 contains all 52 search and replace statements for legal form AG to be removed at the end of the name, with the keyword containing the spelling variation identified as legal form, the number of occurrences of the spelling variation in all names, the harmonized legal form, and remarks on how to replace the legal form. Appendix 2 contains all search and replace statements for all legal forms to be removed at the end of a name.

**Table 15: Search and replace statements for legal form AG to be removed at end of name**

| KEYWORD | NBR | LEGAL FORM | REMARKS |
|---|---|---|---|
| " AG" | 6,156 | AG | Remove |
| " AKTIENGESELLSCHAFT" | 1,141 | AG | Remove |
| " A.G." | 246 | AG | Remove |
| " AG." | 104 | AG | Remove |
| ", AG" | 96 | AG | Remove |
| " CO. AG" | 67 | AG | Replace with " COMPANY" |
| " AG & CO." | 43 | AG | Replace with " & COMPANY" |
| ", A.G." | 30 | AG | Remove |
| " CIE AG" | 13 | AG | Replace with " COMPANY" |
| " CIE. AG" | 11 | AG | Replace with " COMPANY" |
| " CO. AKTIENGESELLSCHAFT" | 9 | AG | Replace with " COMPANY" |
| " AG & CO" | 8 | AG | Replace with " & COMPANY" |
| " AKTIENGESELLSCHAFT & CO." | 7 | AG | Replace with " & COMPANY" |
| ", AKTIENGESELLSCHAFT" | 6 | AG | Remove |
| " CO AG" | 5 | AG | Replace with " COMPANY" |
| " + CO. AG" | 5 | AG | Replace with " & COMPANY" |
| " AKTIENGESELLSCHAFT AG" | 5 | AG | Remove |
| ", AG." | 5 | AG | Remove |

| | | | |
|---|---|---|---|
| " CO., AG" | 4 | AG | Replace with " COMPANY" |
| " GMBH & CO. AG" | 4 | AG | Replace with " & COMPANY" |
| " (AG)" | 4 | AG | Remove |
| " CO. AG." | 4 | AG | Replace with " COMPANY" |
| " AG + CO" | 3 | AG | Replace with " & COMPANY" |
| " AG AKTIENGESELLSCHAFT" | 3 | AG | Remove |
| " AG CO." | 3 | AG | Replace with " COMPANY" |
| " A.-G." | 3 | AG | Remove |
| " AKTIEN-GESELLSCHAFT" | 3 | AG | Remove |
| " CO AKTIENGESELLSCHAFT" | 3 | AG | Replace with " COMPANY" |
| " CO., AKTIENGESELLSCHAFT" | 2 | AG | Replace with " COMPANY" |
| " + CIE AG" | 2 | AG | Replace with " & COMPANY" |
| " A/G" | 2 | AG | Remove |
| " GMBH & CO AG" | 2 | AG | Replace with " & COMPANY" |
| " AG + CO." | 2 | AG | Replace with " & COMPANY" |
| " CIE. AG." | 2 | AG | Replace with " COMPANY" |
| " CIE AKTIENGESELLSCHAFT" | 2 | AG | Replace with " COMPAGNIE" |
| " AG & CO AG" | 1 | AG | Replace with " & COMPANY" |
| " AKTIENGESELLSCHAFT, AG" | 1 | AG | Remove |
| " + CO AKTIENGESELLSCHAFT" | 1 | AG | Replace with " & COMPANY" |
| ", AG & CO." | 1 | AG | Replace with " & COMPANY" |
| " + CO. AKTIENGESELLSCHAFT" | 1 | AG | Replace with " & COMPANY" |
| " A.G. & CO." | 1 | AG | Replace with " & COMPANY" |
| " A.G. AKTIENGESELLSCHAFT" | 1 | AG | Remove |
| ", A.G" | 1 | AG | Remove |
| " CO., A.G." | 1 | AG | Replace with " COMPANY" |
| " AKTIENGESELLSCHAFT UND CO." | 1 | AG | Replace with " & COMPANY" |
| " CIE. AKTIENGESELLSCHAFT" | 1 | AG | Replace with " COMPAGNIE" |
| " AKTIENGESELL-SCHAFT" | 1 | AG | Remove |
| " CIE. A.-G." | 1 | AG | Replace with " COMPANY" |
| " AG+ CO." | 1 | AG | Replace with " & COMPANY" |
| " CO. A.G." | 1 | AG | Replace with " COMPANY" |
| " AG AND CO." | 1 | AG | Replace with " & COMPANY" |
| " GMBH & CO., AG" | 1 | AG | Replace with " & COMPANY" |

The keyword to look for contains the identified spelling variation. For every spelling variation, the number of occurrences at the end of the string is indicated.

The legal from contains the harmonized legal form ("AG" in this case). In general, abbreviations are used because several legal forms of different countries might have the same abbreviation (e.g. Société Anonyme and Sociedad Anónima), making it not always possible at the level of the name to identify the real underlying legal form. Additional address (country) information can be used to identify the exact legal form, although this can lead to misleading results for multinational companies.

Mostly, the legal form indication can be removed at the end of the name ("Remove" in field REMARK) but, in some cases, some additional harmonization is carried out (e.g., in a name ending with " CO. A.G.", the real legal form " A.G." will be removed and the element " CO." will be replaced with " COMPANY"). This harmonization is carried out for words that are not really part of the legal form but appear in combination with it, and that can be harmonized immediately.

Table 16 contains some examples of word variations commonly appearing before a legal form indication and their harmonized equivalent.

**Table 16: Variations and harmonized equivalent of words commonly appearing before legal form indications**

| VARIATIONS IN NAME | HARMONIZED EQUIVALENT |
|---|---|
| CO, CO., … | "COMPANY" |
| AND CO, AND COMPANY, … | "& COMPANY" |
| CIE, CIE., … | "COMPAGNIE" |
| ET CIE, … | "& COMPAGNIE" |
| INTL, INT'L | "INTERNATIONAL" |
| CORP, CORP., … | "CORPORATION" |
| MFG, MFG., … | "MANUFACTURING" |

Some legal form indications are not removed from the name because they are, to all intents and purposes, part of the name and removing them could make the underlying name less comprehensible. A typical example is the German Kommanditgesellschaft, abbreviated as "KG". The part "gesellschaft" is part of the full name in a significant number of cases, as in "ESPE STIFTUNG & CO. PRODUKTION - UND VERTRIEBS KG". To prevent mutilation of these kinds of name, "KG" will not be removed from the name. In one of the next steps - common company word removal - "KG" will be removed to obtain a reduced name that can be used for searching related names.

Table 15 also shows that the order of removing and harmonizing legal form indications is important. It is clear that " + CO. AG" should be replaced with " & COMPANY" before removing " AG".

All 1,060 spelling variations of legal forms occurring at the end of names were converted in search and replace statements or rules, as listed in Table 15. All these rules and spelling variations were validated. For every rule or spelling variation, all names containing the spelling variation were scanned to be certain that the rule only affects actual legal form indications. If the number of occurrences was higher than 500 (47 of the 1,060 spelling variations), only a sample of 500 names was checked.

A rule or search and replace statement is withheld only if more than 99% of found spelling variations are actually legal form indications. As the vast majority of rules resulted in 0 mistakes, the overall accuracy is greater than 99%.

The full list of all search and replace statements for legal forms to be removed at the end of a name can be found in Appendix 2.

It has to be stressed that the objective is not to maximize the total number of matches (at the cost of introducing mismatches) but to minimize the number of mismatches given a reasonable number of matches.

This means that a considerable number of legal form indications will still be present in the names after legal form removal. On the one hand, only legal forms that were identified on the basis of the top 40 occurring last words were removed and harmonized, leaving a substantial number of legal forms unchanged. On the other hand, not all spelling variations of identified legal form indications were removed because some of the occurrences might have nothing to do with legal form indications but may be mere coincidences.

In addition to the real legal form removal and harmonization, some other words, commonly used in a company context, appearing as last words were also harmonized:

- "CO" was harmonized to "COMPANY". Preceding "+", "AND", "U", or "UND" were harmonized to "&"
- "AND" preceding "COMPANY" was harmonized to "&"
- " CORP" was harmonized to " CORPORATION"
- " E C." was harmonized to " & COMPANY"
- " & C." and " & C" were harmonized to " & COMPANY"

At this stage, only legal form indications appearing at the end of a name were removed and harmonized. In some countries, legal form indications can also appear in front of company names. The approach described above for last words can also be used for first words.

Table 17 contains the top 50 occurring first words after cleanup, together with the number of names containing the word as a first word, the cumulative number of names for this word and all higher ranked words, and the percentage of the cumulative number of names compared to the total number of names (443,722). First words are identified on the basis of the first occurrence of a space in a name, then all non-(A-Z) and non-(0-9) characters are removed resulting in a cleaned version of the first word. Appendix 4 contains the list of the top 200 occurring first words.

**Table 17: Top 50 occurring first words**

| FIRST WORD (CLEANED) | NBR NAMES | CUM | % | FIRST WORD (CLEANED) | NBR NAMES | CUM | % |
|---|---|---|---|---|---|---|---|

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | THE | 5489 | 5489 | 1.2 | 26 | SMITH | 433 | 28532 | 6.4 |
| 2 | | 5385 | 10874 | 2.5 | 27 | R | 418 | 28950 | 6.5 |
| 3 | SOCIETE | 2477 | 13351 | 3.0 | 28 | C | 416 | 29366 | 6.6 |
| 4 | KABUSHIKI | 1293 | 14644 | 3.3 | 29 | COMPAGNIE | 386 | 29752 | 6.7 |
| 5 | ADVANCED | 1076 | 15720 | 3.5 | 30 | APPLIED | 386 | 30138 | 6.8 |
| 6 | AMERICAN | 943 | 16663 | 3.8 | 31 | SIEMENS | 380 | 30518 | 6.9 |
| 7 | INTERNATIONAL | 936 | 17599 | 4.0 | 32 | KIM | 370 | 30888 | 7.0 |
| 8 | VAN | 923 | 18522 | 4.2 | 33 | DR | 364 | 31252 | 7.0 |
| 9 | NIPPON | 861 | 19383 | 4.4 | 34 | M | 364 | 31616 | 7.1 |
| 10 | DE | 767 | 20150 | 4.5 | 35 | AB | 351 | 31967 | 7.2 |
| 11 | UNIVERSITY | 686 | 20836 | 4.7 | 36 | CENTRE | 345 | 32312 | 7.3 |
| 12 | INSTITUT | 615 | 21451 | 4.8 | 37 | FUJI | 337 | 32649 | 7.4 |
| 13 | HITACHI | 604 | 22055 | 5.0 | 38 | H | 335 | 32984 | 7.4 |
| 14 | NATIONAL | 595 | 22650 | 5.1 | 39 | TOKYO | 325 | 33309 | 7.5 |
| 15 | JAPAN | 544 | 23194 | 5.2 | 40 | E | 322 | 33631 | 7.6 |
| 16 | UNITED | 540 | 23734 | 5.3 | 41 | US | 313 | 33944 | 7.6 |
| 17 | J | 532 | 24266 | 5.5 | 42 | ABB | 311 | 34255 | 7.7 |
| 18 | A | 519 | 24785 | 5.6 | 43 | B | 298 | 34553 | 7.8 |
| 19 | GENERAL | 515 | 25300 | 5.7 | 44 | CHEN | 296 | 34849 | 7.9 |
| 20 | NEW | 499 | 25799 | 5.8 | 45 | G | 289 | 35138 | 7.9 |
| 21 | LEE | 485 | 26284 | 5.9 | 46 | SA | 288 | 35426 | 8.0 |
| 22 | FIRMA | 472 | 26756 | 6.0 | 47 | S | 287 | 35713 | 8.0 |
| 23 | JOHNSON | 459 | 27215 | 6.1 | 48 | SUMITOMO | 283 | 35996 | 8.1 |
| 24 | MITSUBISHI | 449 | 27664 | 6.2 | 49 | W | 281 | 36277 | 8.2 |
| 25 | ETABLISSEMENTS | 435 | 28099 | 6.3 | 50 | UNIVERSAL | 280 | 36557 | 8.2 |

Table 17 shows a completely different picture to Table 13. The number of occurrences is far lower for top occurring first words than for top occurring last words (5,489 for "THE" compared to 74,949 for "INC"), as is the cumulative share of top 50 first words compared to the top 50 last words (8.2% compared to 65%). This means that the variety of words at the beginning of organization names is far greater than the variety of words at the end of organization names.

Table 17 reveals that only "SOCIETE" and "KABUSHIKI" are worth looking at more closely.

"KABUSHIKI" in itself is not a legal form but "KABUSHIKI HAISHA" is; it can, therefore, be removed at the beginning of a name.

"SOCIETE" in itself is not a legal form but "SOCIETE ANONYME" or "SOCIETE A RESPONSABILITE LIMITEE" is. The problem is that these legal form indications can be an integral part of the name (as in "SOCIETÉ ANONYME D'ECONOMIE MIXTE COMMUNAUTAIRE DE GESTION" or "SOCIETE ANONYME DES ETABLISSEMENTS PIERRE ROCH") and removing them can mutilate names and make them hard to recognize.

As is the case of "KG" (Kommanditgesellschaft), "SOCIETE" will not be removed in this step, even as part of "SOCIETE ANONYME" or "SOCIETE A RESPONSABILITE LIMITEE", but will be removed in one of the subsequent steps - common company word removal - to obtain a reduced name that can be used for searching related names.

Finally, legal forms can also appear in the middle of names. A quick examination based on a full text index revealed that occurrences are low. The expected impact seems not to justify the efforts needed for an in-depth analysis of occurrences, spelling variations and validation.

The only legal form indication clearly appearing frequently in the middle of company names is "GMBH".

Table 18 contains all 17 search and replace statements for legal form GMBH to be removed anywhere in the name, with the keyword containing the spelling variation identified as a legal form, the number of occurrences of the spelling variation in all names, the harmonized legal form, and remarks on how to remove the legal form.

**Table 18: Search and replace statements for legal form GMBH to be removed anywhere in name**

| KEYWORD | NBR | LEGAL | REMARKS |
|---|---|---|---|
| " GMBH & CO. K.G. " | 1 | GMBH | Replace with " & COMPANY " |
| " GMBH & CO. KG. " | 8 | GMBH | Replace with " & COMPANY " |

| | | | |
|---|---|---|---|
| " GMBH & CO. KG " | 191 | GMBH | Replace with " & COMPANY " |
| " GMBH & CO.K.G. " | 0 | GMBH | Replace with " & COMPANY " |
| " GMBH & CO.KG " | 15 | GMBH | Replace with " & COMPANY " |
| " GMBH & CO KG " | 18 | GMBH | Replace with " & COMPANY " |
| " GMBH + CO. KG " | 13 | GMBH | Replace with " & COMPANY " |
| " GMBH & CO. " | 612 | GMBH | Replace with " & COMPANY " |
| " GMBH & CO " | 65 | GMBH | Replace with " & COMPANY " |
| " GMBH & CO.," | 77 | GMBH | Replace with " & COMPANY " |
| " GMBH & CO," | 8 | GMBH | Replace with " & COMPANY " |
| " GMBH + CO. " | 28 | GMBH | Replace with " & COMPANY " |
| " GMBH + CO " | 6 | GMBH | Replace with " & COMPANY " |
| " GMBH + CO.," | 1 | GMBH | Replace with " & COMPANY " |
| " GMBH + CO," | 1 | GMBH | Replace with " & COMPANY " |
| " GMBH," | 173 | GMBH | Remove |
| " GMBH " | 1,648 | GMBH | Remove |

### Implementation

The more complex the analysis (identification of legal forms, identification of spelling variations, validation), the simpler the implementation.

All identified and validated spelling variations of legal form indications are transferred to search and replace statements or rules as in Table 15 and Table 18. This results in 1,060 rules or statements to remove legal form indications at the end of names, one rule or statement to remove legal form indications at the beginning of names, and 17 rules or statements to remove legal form indications anywhere in the name.

Every statement or rule contains the spelling variation to identify and the harmonized string to substitute. In most cases, legal form indications are simply removed and not replaced with anything; replacement is used if legal form indication is preceded or followed by general company words that can be harmonized (e.g. harmonize " + CO." to " & COMPANY").

Every statement or rule also includes the harmonized legal form. This is used to update a new field with the harmonized legal form. The legal form indications are not deleted completely but instead removed from the name field and moved in a harmonized format to a different field.

All identified and validated occurrences of legal form indications are removed by executing a program that reads the search and replace statements or rules, and executes an update query on the data to replace the given keyword (spelling variation of legal form indication) with a given string (mostly replaced with nothing to simply remove the legal form indication) while, at the same time, updating a new field to contain the harmonized legal form of the organization.

The search and replace statements were executed in three groups: firstly, a group of 1,060 statements to remove legal forms at the end of a name (see Appendix 2); then a group of 1 statement to remove legal forms at the beginning of a name (remove 1216 occurrences of "KABUSHIKI KAISHA" at the beginning of a name); and finally, a group of 17 statements to remove legal forms anywhere in a name (see Table 18).

In a group, all search and replace statements are executed in a singular and not a cumulative approach. This means that if a name is updated because of a search and replace statement, it cannot be updated again in a subsequent search and replace statement. This is to prevent a cascade of replacements within one name leading to unexpected results (with consequent difficulties for checking and validating). The list of search and replace statements is constructed bearing this important implementation consideration in mind (e.g. first replace " + CO. AG" with " & COMPANY" before removing " AG").

If a name contains a legal form indication at the beginning and the end of a name, or anywhere in the name, only the legal form indication occurring at the end of the name is harmonized and moved to a different field.

Not all search and replace functions are concerned with legal form indications removal. Some words, commonly used in a company context, appearing as last words were also harmonized, such as "CO" and "CORP" (see analysis for more details).

As the replacements and removals in the search and replace statements can lead to names ending with irregular punctuation characters, all occurrences of "-"; ";"; ":"; "," and "&" are removed at the end of a name by executing an update query on the data.

Finally, as the replacements and removals in the search and replace statements can also lead to leading or trailing spaces, names have to be checked for and trimmed of leading and trailing spaces after removal of legal form indications.

### Result

Legal form indications have been removed and harmonized at the end of names in 221,498 names, at the beginning of names in 1,216 names, and anywhere in the name in 2,865 names.

Moreover, words, commonly used in a company context, appearing as last words were harmonized in 9,150 cases.

Since not all legal form indications have been identified, nor all spelling variations of identified legal form indications have been identified and added to the list of search and replace statements (because they would cause too many false matches), a significant number of names will still contain legal form indications. However, the vast majority of legal form indications are removed, with accuracy well above 99%.

### Impact

From 437,336 unique names to 392,226 unique names, an additional reduction of 45,110 names, or a total reduction of 51,496 names (11.6%).

## 2.2 Common company word removal

### Description

In addition to the legal form indication that can be removed as it is not really part of the name, there are some other words commonly used in a company context that are not really distinctive elements of a company name. Such words include "COMPANY", "CORPORATION", "GESELLSHAFT" and "SOCIETE".

The idea is that if two names are found that are completely identical except for these words, the underlying organization name will be the same and these words can be removed.

Examples include "3COM" and "3COM CORPORATION", "AMIC" and "AMIC COMPANY", "BAUR SPEZIALTIEFBAU" and "BAUR SPEZIALTIEFBAU GESELLSCHAFT", "SOCIETE NOVATEC" and "NOVATEC".

In addition, legal forms identified but not removed in the previous step – legal form indication treatment - are removed at this stage. Such legal forms as "KG" were not removed previously because they are, to all intents and purposes, part of the name and removing them could make the underlying name less comprehensible.

Common company word removal can mutilate organization names and make them less understandable. However, the idea is not to use these common company word removal names as final harmonized names but as some kind of technical search name that can be used to identify name variations in the same organization.

### Analysis

Common company words that can be removed were identified by using the last word index and first word index employed in the previous step – legal form indication treatment - and a full text index of the organization names.

Firstly, the occurrence of "CORPORATION", "COMPANY", "KG" and "GESELLSCHAFT" at the end of names, identified but not always completely removed in the previous step – legal form indication treatment - were analyzed by manually scanning for all spelling variations.

Table 19 contains all spelling variations of all common company words that can be deleted if they appear at the end of a name.

**Table 19: Common company words to be removed at the end of a name**

| KEYWORD | NBR |
|---|---|
| "CORPORATION" | 23,134 |
| "CORP" | 102 |

| | |
|---|---|
| "AND COMPANY" | 120 |
| "& COMPANY" | 10,909 |
| "COMPANY" | 30,946 |
| " KG" | 1,078 |
| "GESELLSCHAFT" | 1,863 |

Next, the occurrence of "SOCIETE" at the beginning of names, identified but not always completely removed in the previous step – legal form indication treatment - was analyzed by manually scanning for all spelling variations.

Table 20 contains all spelling variations of this word that can be deleted if they appear at the beginning of a name.

**Table 20: Common company words to be removed at the beginning of a name**

| KEYWORD | NBR |
|---|---|
| "SOCIETE A RESPONSABILITE LIMITEE DITE" | 20 |
| "SOCIETE A RESPONSABILITE LIMITEE" | 19 |
| "SOCIETE ANONYME DITE" | 130 |
| "SOCIETE ANONYME DES " | 40 |
| "SOCIETE ANONYME DE " | 23 |
| "SOCIETE ANONYME D'" | 14 |
| "SOCIETE ANONYME" | 110 |
| "SOCIETE CIVILE DES " | 6 |
| "SOCIETE CIVILE DE " | 12 |
| "SOCIETE CIVILE D'" | 18 |
| "SOCIETE CIVILE " | 52 |
| "SOCIETE DITE" | 60 |
| "SOCIETE DES " | 104 |
| "SOCIETE DE " | 260 |
| "SOCIETE D'" | 310 |
| "SOCIETE " | 1,285 |

Finally, the full index was scanned to look for other words that are not distinctive elements in company names. Again, all kinds of variations of "CORPORATION", "COMPANY", "GESELLSCHAFT" and "SOCIETE" were identified that can be removed anywhere in a name. In addition, legal form indications "INC" and "AG" still occur frequently and can also be removed anywhere in a name.

Table 21 contains all spelling variations of all these words that can be deleted if they appear anywhere in a name.

**Table 21: Common company words to be removed anywhere in a name**

| KEYWORD | NBR |
|---|---|
| " AND CO " | 0 |
| " AND CO." | 4 |
| " AND CO," | 0 |
| " & CO " | 25 |
| " & CO." | 415 |
| " & CO," | 3 |
| " CO " | 23 |
| " CO." | 859 |
| " CO," | 8 |
| " GESELLSCHAFT " | 1,510 |
| " SOCIETE " | 922 |
| " CORPORATION " | 802 |
| " INC." | 817 |
| " INC," | 8 |
| " COMPANY " | 1,619 |
| " AG " | 559 |
| " AG," | 80 |
| " AG." | 8 |

### Implementation

Implementation is straightforward. All identified spelling variations in Table 19, Table 20 and Table 21 are transferred to search and replace statements or rules as in the previous step – legal form indication treatment.

All identified and validated occurrences of common company words were removed by executing a program that reads the search and replace statements or rules, and executes an update query on the data to replace the given keyword (spelling variation of common company word) with a given string (replace with nothing to simply remove the common company word), while at the same time updating a new field to contain the found spelling variation.

### Result

Common company words have been removed at the end of names in 68,152 names, at the beginning of names in 2,463 names, and anywhere in the name in 7,662 names.

Not all common words that are not distinctive elements in names are removed; only the most commonly used ones are identified by using the last word index, first word index and full text index. A more in-depth analysis of the indexes could reveal additional words safe to remove.

### Impact

From 392,226 unique names to 385,771 unique names, an additional reduction of 6,455 names, or a total reduction of 57,951 names (13.1%).

## 2.3 Spelling variation harmonization

### Description

One of the causes of name variations is spelling variation (mistakes, typographical errors, etc.). Identification of word similarities with approximate string searching (for example, based on Levenshtein distance or edit distance) can be used to identify spelling variations. The problem is that it is not possible to validate name variations in proper names.

For example, "AMTECH" and "IMTECH" have a Levenshtein distance of 1 but is it possible to combine them into one organization name?

However, spelling, language and grammatical variations are identifiable in the case of plain English words or other languages.

For example, "SYSTEM", "SYSTEMS", "SYSTEMEN", "SYSTEMES" can all be harmonized to "SYSTEM" or "SYSTEMS".

Spelling variation harmonization can mutilate organization names and make them less comprehensible. However, the idea is not to use these spelling-variation harmonized names as final harmonized names but as some kind of technical search name that can be used to identify name variations of the same organization.

### Analysis

Spelling variations that can be harmonized were identified by using a full text index of the organization names.

By sorting the index on the number of occurrences, most commonly used words can be identified. Then, by sorting the index alphabetically, variations of those commonly used words can be identified.

Table 22 contains spelling variations of words that can be harmonized.

**Table 22: Spelling variations and their harmonized equivalent**

| KEYWORD | NBR | REMARKS |
|---|---|---|
| "SYSTEMEN" | 48 | "SYSTEM" |
| "SYSTEMES" | 164 | "SYSTEM" |
| "SYSTEME" | 1,140 | "SYSTEM" |
| "SYSTEMS" | 10,104 | "SYSTEM" |
| "INTERNATIONALE" | 109 | "INTERNATIONAL" |

| | | |
|---|---|---|
| "TECHNOLOGIES" | 7,587 | "TECHNOLOGY" |
| "TECHNOLOGIEN" | 61 | "TECHNOLOGY" |
| "TECHNOLOGIE" | 705 | "TECHNOLOGY" |
| "INDUSTRIELLES" | 112 | "INDUSTRIEL" |
| "INDUSTRIELLE" | 415 | "INDUSTRIEL" |
| "INDUSTRIELE" | 16 | "INDUSTRIEL" |
| "INDUSTRIES" | 6,095 | "INDUSTRY" |
| "INDUSTRIELS" | 71 | "INDUSTRIEL" |
| "INSTITUT" | 3,753 | "INSTITUTE" |
| "SERVICES" | 2,181 | "SERVICE" |
| "ELECTRONICS" | 2,742 | "ELECTRONIC" |
| "ENTERPRISES" | 1,622 | "ENTERPRISE" |
| "DESIGNS" | 358 | "DESIGN" |
| "CHEMICALS" | 899 | "CHEMICAL" |
| "HOLDINGS" | 1,457 | "HOLDING" |
| "LABORATORIES" | 1,373 | "LABORATORY" |
| "COMMUNICATIONS" | 1,521 | "COMMUNICATION" |
| "INSTRUMENTS" | 992 | "INSTRUMENT" |
| "PLASTICS" | 959 | "PLASTIC" |
| "MACHINES" | 388 | "MACHINE" |
| "SCIENCES" | 843 | "SCIENCE" |

### Implementation

Implementation is again very simple and straightforward. All identified spelling variations in Table 22 are transferred into search and replace statements or rules as in the previous steps.

All identified and validated occurrences of spelling variations are harmonized by executing a program that reads the search and replace statements or rules, and executes an update query on the data to replace the given keyword (spelling variation) with a given string (harmonized word), while at the same time updating a new field to contain the found spelling variation.

### Result

Spelling variations have been harmonized in 45,715 names.

By and large, not all spelling and language variations have been harmonized; only the most commonly used ones were identified by using the full text index. A more in-depth analysis of the indexes could reveal additional words safe to harmonize.

### Impact

From 385,771 unique names to 384,235 unique names, an additional reduction of 1,536 names, or a total reduction of 59,487 names (13.4%).

## 2.4 Condensing

### Description

After implementing all previous cleaning steps, a significant number of variations are still present because of alternative spellings caused by separation or punctuation characters and all other kinds of non-alphanumerical characters that are not relevant to identify a name (e.g. "3 COM" and "3COM", and "AAF-MCQUAY", "AAF MCQAY" and "AAF – MCQAY").

### Analysis

Condensing names by simply removing all non-alphanumerical characters and spaces and leaving only letters and numbers is enough to reduce many variations without introducing mismatches.

### Implementation

All non-alphanumerical characters are removed by executing a program that reads the names character by character, removing all characters not in the range of a-z, A-Z and 0-9.

### Result

383,707 names contain spaces or non-alphanumerical characters and have been condensed.

### Impact

From 384,235 unique names to 365,866 unique names, an additional reduction of 18,369 names, or a total reduction of 77,856 names (17.5%).

## 2.5 Umlaut harmonization

### Description

As described in a previous step - replace accented characters - German characters with a diacritic mark ('umlaut': "ä", "ö", "ü") - cause spelling variations because words containing these characters can occur in three guises, one with an umlaut (e.g. "für"), one with the alternative spelling without an umlaut but with an additional "e" (e.g. "fuer"), and a simplified form without an umlaut and without an additional "e" (e.g. "fur").

Since all of these spelling variations appear in the organization names, simply replacing all characters containing an umlaut with their simple underlying equivalent without an umlaut and without an additional "e", as in the earlier cleaning step, will not match all equivalent names.

This additional step will try to match the spelling variant without an umlaut but with an additional "e" with the other spelling variations.

Other languages such as Hungarian also suffer from this problem but, in this step, emphasis is placed on the German umlaut and its equivalent with an additional "e".

### Analysis

Since all three variations appear in organization names (sometimes more than one variation in a name, e.g. "PATENT-TREUHEND-GESELLSCHAFT FUER ELEKTRISCHE GLÜHLAMPEN MBH"), no straightforward solution is available.

Given the former example, creating two variations of all names with umlauts, one without an umlaut but with an additional "e", and one without an umlaut and without an additional "e", will not work because all kinds of combinations can appear in one name. Even if "PATENT-TREUHEND-GESELLSCHAFT FUER ELEKTRISCHE GLÜHLAMPEN MBH" could be harmonized both to "PATENT-TREUHEND-GESELLSCHAFT FUR ELEKTRISCHE GLUHLAMPEN MBH" and "PATENT-TREUHEND-GESELLSCHAFT FUER ELEKTRISCHE GLUEHLAMPEN MBH", the following variation "PATENT-TREUHEND-GESELLSCHAFT FUR ELEKTRISCHE GLUEHLAMPEN MBH" would not be matched.

Therefore, not only names containing umlauts have to be harmonized but all names will have to be scanned for possible matches with a name containing an umlaut.

Simply adding an "e" to, or removing it from, all occurring "a", "o" or "u" leads to many mismatches, especially in the case of proper names containing "a", "o" or "u".

To eliminate these mismatches, only groups of matched names with at least one name originally containing at least one umlaut are retained for the name harmonization. However, this additional step to maintain accuracy greatly reduces the number of matches.

### Implementation

Firstly, all occurrences of "AE", "OE" and "UE" are replaced with "A", "O" and "U" respectively in all names (also in names originally containing no umlauts) by executing a series of update queries on the data.

Next, all occurrences of "A", "O" and "U" are again replaced with "AE", "OE", "UE" respectively in all names (also in names originally containing no umlauts) by executing a series of update queries on the data.

Next, all names originally containing an umlaut are marked by executing an update query on the data.

Next, all names having a preliminary umlaut harmonized name (first removing "E" and next adding "E" from and to "A", "O" and "U") that is equal to a preliminary umlaut harmonized name

marked in the previous step as a name originally containing an umlaut are also marked by executing an update query on the data.

Finally, all preliminary umlaut harmonized names not marked in the previous two steps are reverted to the previous cleaned name after condensing by executing an update query on the data.

### Result

Umlauts have been harmonized in 9,443 names.

By and large, not all umlaut variations have been harmonized because refining the method to increase the number of matches tends to increase the number of mismatches quite substantially.

The method presented here is very safe (100% correct matches) but it could well be improved to cover more names.

### Impact

From 365,866 unique names to 365,564 unique names, an additional reduction of 302 names, or a total reduction of 78,158 names (17.6%).

## 2.6 Cleaned name

### Final result

The final cleaned name is the name after character cleaning, punctuation cleaning, legal form indication treatment, common company word removal, spelling variation harmonization, condensing and umlaut harmonization.

During cleaning, the original name can become heavily mutilated and unrecognizable (e.g., from "" NEUSON" -ÖLFELDSCHIEBER GESELLSCHAFT M.B.H." to "NEUESOENOELFELDSCHIEBER"). In this stage, the cleaned name is only usable to identify matching names. In a subsequent step, the cleaned name will be converted back to a more readable and usable name closer to the original.

In total, all 443,722 names have been affected by one of the cleaning or harmonization steps.

### Final impact

All cleaning and harmonization steps resulted in a reduction from 443,722 unique original names to 365,564 unique cleaned names, a reduction of 78,158 names or 17.6%.

Table 23 contains an overview of the impact of every step, with the number of unique names before and after the particular cleaning and harmonization step, the reduction in the number of names, the cumulative reduction in the number of names for the particular step and all previous steps, the relative reduction compared to the total number of original unique names, and the cumulative relative reduction compared to the total number of original unique names.

**Table 23: step by step results of cleaning and harmonization**

| STEP | FROM | TO | REDUCTION | REDUCTION CUM | % | % CUM |
|---|---|---|---|---|---|---|
| Character cleaning | 443,722 | 438,366 | 5,356 | 5,356 | 1.2 | 1.2 |
| Punctuation cleaning | 438,366 | 437,336 | 1,030 | 6,386 | 0.2 | 1.4 |
| Legal form removal | 437,336 | 392,226 | 45,110 | 51,496 | 10.2 | 11.6 |
| Common company word removal | 392,226 | 385,771 | 6,455 | 57,951 | 1.5 | 13.1 |
| Spelling variation harmonization | 385,771 | 384,235 | 1,536 | 59,487 | 0.3 | 13.4 |
| Condensing | 384,235 | 365,866 | 18,369 | 77,856 | 4.1 | 17.5 |
| Umlaut harmonization | 365,866 | 365,564 | 302 | 78,158 | 0.1 | 17.6 |

Legal form removal and condensing are, by far, the most important steps. This does not mean that all other steps can be neglected, as these steps prepare the data for subsequent steps. Even if the impact of a particular step is low, the results of the step can greatly improve the impact of the steps that follow.

# 3 HARMONIZATION RESULTS

As the final cleaned name can be heavily mutilated and hard to recognize, a final, more readable, harmonized name is added.

For every cleaned name, the legal-form-removed version of the original name having the most patents assigned to it was taken as the final harmonized name. This name is already harmonized to a degree, but is still completely readable and recognizable.

If the legal-form-removed name is empty (e.g. original name is "SOCIETE ANONYME"), the original name is taken as the harmonized name.

If the end user does not want to combine harmonized names with different legal forms, the combination of the final harmonized name and the harmonized legal form as created by the legal form removal can be used as the final name.

The overview in Table 23 already contains the reduction in unique names but this is only a partial picture of the real impact of the name cleaning and harmonization procedure.

Important aspects of name cleaning and harmonization also concern the extent to which:

- original names are matched to harmonized names;
- additional patents are assigned to harmonized names;
- patent distribution amongst patentees has changed;
- patent ranking of patentees has changed.

## 3.1 Original names matched to harmonized names

Of the final 365,566 unique harmonized names, 49,449 (13.5%) names match more than one original name, ranging from 2 to 51 names.

The number of unique harmonized names (365,566) is slightly different from the number of unique cleaned names (365,564) because original patentee names containing only legal form indications or common company words are reduced to empty strings during the cleaning process. The original patentee name is taken as the harmonized name for those three names ("SOCIETE A RESPONSABILITE LIMITEE", "SOCIETE ANONYME" and "SOCIETE ANONYME DITE"), resulting in two additional unique harmonized names (the number of unique cleaned names minus 1 for the one unique empty name, plus 3 for the underlying three original names).

Table 24 contains all 51 identified name variations (original patentee names) of the two harmonized names having the most matched names, "E.I. DU PONT DE NEMOURS & COMPANY" and "SGS-THOMSON MICROELECTRONICS".

**Table 24: Matched name variations of "E.I. DU PONT DE NEMOURS & COMPANY" and "SGS-THOMSON MICROELECTRONICS"**

| E.I. DU PONT DE NEMOURS & COMPANY | SGS-THOMSON MICROELECTRONICS |
|---|---|
| E I DU PONT DE NEMOURS AND COMPANY | S.G.S. THOMSON MICROELECTRONICS S.R.L. |
| E I DUPONT DE NEMOURS AND COMPANY | S.G.S. THOMSON MICROELECTRONICS, S.R.L. |
| E I. DU PONT DE NEMOURS AND COMPANY | S.G.S.-THOMSON MICROELECTRONICS S.R.L. |
| E. .I DU PONT DE NEMOURS AND COMPANY | SGS - THOMSON MICROELECTRONICS S.A. |
| E. I DU PONT DE NEMOURS AND COMPANY | SGS - THOMSON MICROELECTRONICS S.R.L. |
| E. I DU PONT DE NEMOURS AND COMPANY. | SGS - THOMSON MICROELECTRONICS, INC. |
| E. I. DU PONT DE NEMOURS | SGS - THOMSON MICROELECTRONICS, S.R.L. |
| E. I. DU PONT DE NEMOURS & CO | SGS THOMSON MICROELECTRONICS S.A. |
| E. I. DU PONT DE NEMOURS & CO. | SGS THOMSON MICROELECTRONICS S.R.L. |
| E. I. DU PONT DE NEMOURS & CO. (INC.) | SGS THOMSON MICROELECTRONICS SA |
| E. I. DU PONT DE NEMOURS & CO., INC. | SGS THOMSON MICROELECTRONICS SRL |
| E. I. DU PONT DE NEMOURS & COMPANY | SGS THOMSON MICROELECTRONICS, INC. |
| E. I. DU PONT DE NEMOURS AND CO. | SGS THOMSON MICROELECTRONICS, S.A. |
| E. I. DU PONT DE NEMOURS AND CO., INC. | SGS- THOMSON MICROELECTRONICS, S.A. |
| E. I. DU PONT DE NEMOURS AND COMPANY | SGS THOMSON MICROELECTRONICS, S.R.L. |

| | |
|---|---|
| E. I. DU PONT DE NEMOURS AND COMPANY, INC. | SGS- THOMSON MICROELECTRONICS<BR>(PTE) LTD. |
| E. I. DU PONT DE NEMOURS AND COMPANY. | SGS THOMSON-MICROELECTRONICS SA |
| E. I. DU PONT DE NEMOURS CO. | SGS-THOMSON MICROELECTRONIC S.A. |
| E. I. DU PONT DE NEMOURS CO., INC. | SGS-THOMSON MICROELECTRONICS |
| E. I. DU PONT DE NEMOURS COMPANY | SGS-THOMSON MICROELECTRONICS GMBH |
| E. I. DU PONT DE NEMOURS COMPANY, INC. | SGS-THOMSON MICROELECTRONICS INC. |
| E. I. DUPONT DE NEMOURS & CO. | SGS-THOMSON MICROELECTRONICS LIMITED |
| E. I. DUPONT DE NEMOURS & COMPANY | SGS-THOMSON MICROELECTRONICS LTD. |
| E. I. DUPONT DE NEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS PTE LTD |
| E. I. DUPONT DENEMOURS & COMPANY | SGS-THOMSON MICROELECTRONICS PTE LTD. |
| E. I. DUPONT DENEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS PTE. LIMITED |
| E. I.DU PONT DE NEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS PTE. LTD. |
| E.I . DU PONT DE NEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS S. R. L. |
| E.I. DU PONT DE NEMOURS & CO. | SGS-THOMSON MICROELECTRONICS S.A |
| E.I. DU PONT DE NEMOURS & CO., INC. | SGS-THOMSON MICROELECTRONICS S.A. |
| E.I. DU PONT DE NEMOURS & COMPANY | SGS-THOMSON MICROELECTRONICS S.P.A. |
| E.I. DU PONT DE NEMOURS & COMPANY, INC | SGS-THOMSON MICROELECTRONICS S.R. L. |
| E.I. DU PONT DE NEMOURS & CO. (INC.) | SGS-THOMSON MICROELECTRONICS S.R.L |
| E.I. DU PONT DE NEMOURS & COMPANY | SGS-THOMSON MICROELECTRONICS S.R.L. |
| E.I. DU PONT DE NEMOURS & COMPANY INC. | SGS--THOMSON MICROELECTRONICS S.R.L. |
| E.I. DU PONT DE NEMOURS & | SGS-THOMSON MICROELECTRONICS SA |
| E.I. DU PONT DE NEMOURS & COMPAY, INC. | SGS-THOMSON MICROELECTRONICS SPA |
| E.I. DU PONT DE NEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS SRL |
| E.I. DU PONT DE NEMOURS<BR>AND CO. | SGS-THOMSON MICROELECTRONICS SRL. |
| E.I. DU PONT DE NEMOURS<BR>AND COMPANY | SGS-THOMSON MICROELECTRONICS, GMBH |
| E.I. DUPONT DE NEMOURS | SGS-THOMSON MICROELECTRONICS, INC |
| E.I. DUPONT DE NEMOURS & CO. | SGS-THOMSON MICROELECTRONICS, INC. |
| E.I. DUPONT DE NEMOURS & COMPANY | SGS-THOMSON MICROELECTRONICS, LTD. |
| E.I. DUPONT DE NEMOURS AND CO. | SGS-THOMSON MICROELECTRONICS, PTE LTD. |
| E.I. DUPONT DE NEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS, S.A. |
| E.I. DUPONT DE NEMOURS AND COMPANY, INC. | SGS-THOMSON MICROELECTRONICS, S.R.L. |
| E.I. DUPONT DENEMOURS & COMPANY | SGS-THOMSON MICROELECTRONICS, S.RL |
| E.I. DUPONT DENEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS, SA |
| E.I.DU PONT DE NEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS, SA. |
| EI DU PONT DE NEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS, SRL |
| EI DUPONT DE NEMOURS AND COMPANY | SGS-THOMSON MICROELECTRONICS,S.R.L. |

For those harmonized names matching more than one original patentee name, on average, 2.6 original patentee names are matched to the harmonized name.

Table 25 contains the distribution of the number of matched names per harmonized name.

**Table 25: Distribution of number of matched names per harmonized name**

| NBR MATCHED NAMES | NBR HARMONIZED NAMES |
|---|---|
| 51 | 2 |
| 50 | 1 |
| 38 | 1 |
| 34 | 2 |
| 33 | 1 |
| 31 | 3 |
| 30 | 2 |
| 29 | 3 |
| 28 | 1 |
| 27 | 2 |
| 26 | 5 |
| 25 | 2 |
| 24 | 6 |
| 23 | 4 |
| 22 | 3 |
| 21 | 3 |
| 20 | 4 |
| 19 | 11 |
| 18 | 7 |
| 17 | 13 |
| 16 | 15 |
| 15 | 20 |
| 14 | 23 |

| | |
|---|---|
| 13 | 37 |
| 12 | 35 |
| 11 | 61 |
| 10 | 82 |
| 9 | 136 |
| 8 | 179 |
| 7 | 337 |
| 6 | 649 |
| 5 | 1,231 |
| 4 | 3,037 |
| 3 | 8,284 |
| 2 | 35,248 |

The distribution of the number of matched names is extremely skewed; the vast majority of harmonized names match only 2 original patentee names.

## 3.2 Additional patents assigned to harmonized names

Although the number of matched original patentee names per harmonized name is an important measure of the performance of the name cleaning and harmonizing procedure, the bottom line is how many additional patents are assigned to harmonized names.

As an example, Table 26 contains the number of patents assigned to all ten original patentee names matched with the harmonized name "3COM CORPORATION".

**Table 26: Number of patents for matched names of "3COM CORPORATION"**

| ORIGINAL PATENTEE NAME | NBR PAT |
|---|---|
| 3COM CORPORATION | 998 |
| 3 COM CORPORATION | 27 |
| 3COM CORP. | 12 |
| 3COM LTD. | 3 |
| 3COM LIMITED | 2 |
| 3 COM CORP. | 1 |
| 3 COM | 1 |
| 3COM CORP | 1 |
| 3COM. CORP. | 1 |
| 3COMCORPORATION | 1 |

After name cleaning and harmonization, 1,047 patents are assigned to "3COM CORPORATION". Without harmonization, these 1,047 patents are scattered over 10 patentee names, ranging from one patent for "3COMCORPORATION" to 998 patents for "3COM CORPORATION".

Without name cleaning and harmonization, a maximum of 998 patents would have been identified for "3COM CORPORATION". This means that name cleaning and harmonization contributes 49 additional patents, or the share of additional found patents is 4.7% of the total number of patents assigned to the harmonized name.

In general, name cleaning and harmonization adds up to 12,704 patents to the harmonized name. The share of additional assigned patents in the total number of patents assigned to the harmonized name ranges from 1% to 86%.

Table 27 contains some examples of harmonized names with the number of matched original patentee names for the particular harmonized name, the total number of patents, the number of patents of the original patentee name having the most assigned patents, the additional number of patents assigned to the particular harmonized name and the share of additional assigned patents in the total number of patents assigned to the particular harmonized name.

**Table 27: Additional assigned patents**

| HARMONIZED NAME | NBR NAMES | NBR PAT | MAX PAT | ADD PAT | SHARE |
|---|---|---|---|---|---|
| INTERNATIONAL BUSINESS MACHINES CORPORATION | 20 | 41,173 | 28,469 | 12,704 | 31% |

| | | | | | |
|---|---|---|---|---|---|
| H. G. WEBER & COMPANY | 7 | 7 | 1 | 6 | 86% |
| E.I. DU PONT DE NEMOURS & COMPANY | 51 | 12,252 | 6,269 | 5,983 | 49% |
| SGS-THOMSON MICROELECTRONICS | 51 | 3,125 | 922 | 2,203 | 70% |

"INTERNATIONAL BUSINESS MACHINES CORPORATION" has the highest number of additional patents (12,704); "H. G. WEBER & COMPANY" has the highest share of additional patents (86%); and "E.I. DU PONT DE NEMOURS & COMPANY" and "SGS-THOMSON MICROELECTRONICS" have the highest number of matched names (51).

Table 27 reveals that there is not necessarily a relation between the number of matched names and the number or share of additional assigned patents. Even a low number of matched names can result in a high number of additional patents.

Table 28 contains the average impact of name cleaning and harmonization for all harmonized names matched to more than one original patentee name.

**Table 28: Average impact of name cleaning and harmonization**

| | |
|---|---|
| Number of matched names | 2.6 |
| Total number of patents | 48.8 |
| Maximum number of patents of original patentee | 41.3 |
| Additional number of assigned patents | 7.6 |
| Share of additional patents | 33.1% |

Although the number of matched names and the additional number of assigned patents seem low on average, the impact on the share of total number of patents is considerable.

Table 29 contains the distribution of the share of additional assigned patents in the total number of patents of all harmonized names matched to more than one original patentee name.

**Table 29: Distribution of share of additional assigned patents**

| SHARE | NBR HARMONIZED NAMES |
|---|---|
| <90% | 23 |
| <80% | 224 |
| <70% | 1,818 |
| <60% | 14,394 |
| <50% | 3,900 |
| <40% | 8,998 |
| <30% | 8,090 |
| <20% | 6,275 |
| <10% | 5,727 |

Table 29 shows that the results of name cleaning and harmonization is not linear for all names but highly case- or name-specific; some patentees gain significantly but for others it hardly matters.

Overall, there seems to be no relation between the number of matched names, the total number of patents, the number of patents of the original patentee name having the most assigned patents, the additional number of patents assigned to the harmonized name, and the share of additional assigned patents in the total number of patents assigned to the harmonized name.

This means that name cleaning and harmonization might be highly relevant. Distribution and ranking can change because not all patentees profit from harmonization to the same extent.

## 3.3 Patent distribution amongst patentees

Name cleaning and harmonization has reduced the number of unique patentee names by 17.6%, from 443,722 to 365,566. This reduction has increased the average number of patents per patentee from 7.2 before harmonization to 8.8 after.

Table 30 contains the distribution of the number of patents per original patentee name and harmonized name.

**Table 30: Distribution of number of patents per original patentee name and harmonized name**

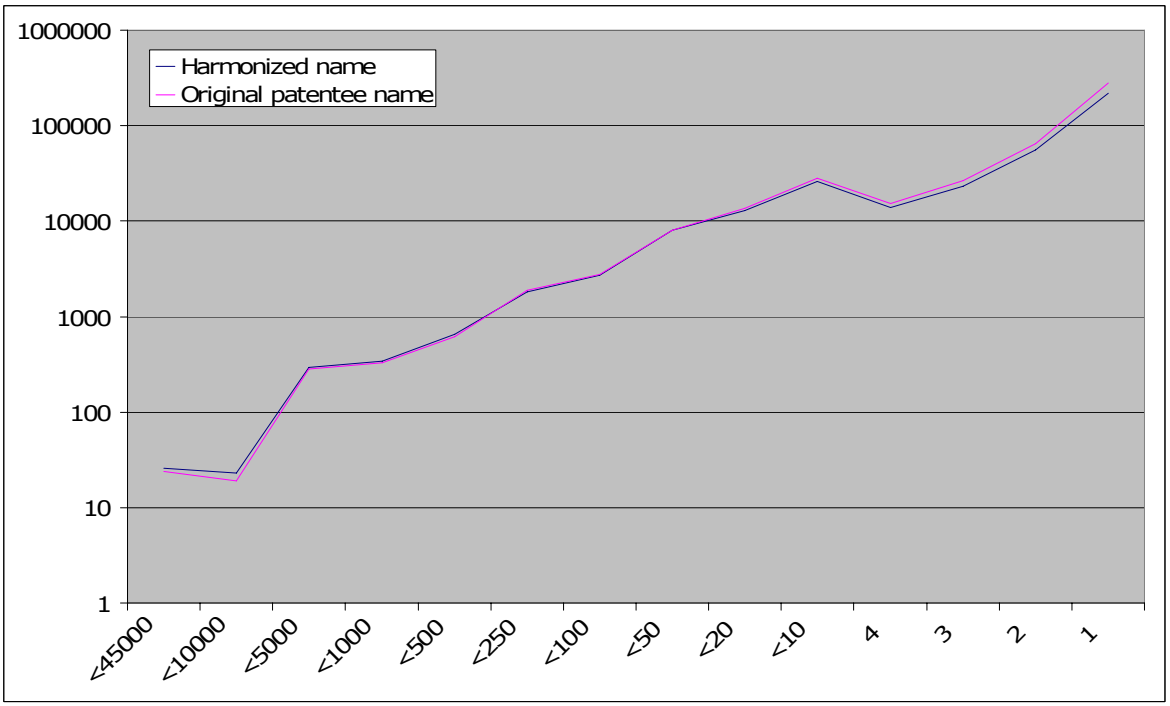
| NBR PAT | NBR PATENTEES BEFORE HARMONIZATION | | NBR PATENTEES AFTER HARMONIZATION | |
|---|---|---|---|---|
| <45,000 | 24 | 0.01% | 26 | 0.01% |
| <10,000 | 19 | 0.01% | 23 | 0.01% |
| <5,000 | 284 | 0.06% | 295 | 0.08% |
| <1,000 | 330 | 0.07% | 342 | 0.09% |
| <500 | 621 | 0.14% | 646 | 0.18% |
| <250 | 1,886 | 0.43% | 1,835 | 0.50% |
| <100 | 2,769 | 0.62% | 2,695 | 0.74% |
| <50 | 8,081 | 1.82% | 7,959 | 2.18% |
| <20 | 13,635 | 3.07% | 12,941 | 3.54% |
| <10 | 28,202 | 6.36% | 25,993 | 7.11% |
| 4 | 15,146 | 3.41% | 13,776 | 3.77% |
| 3 | 26,462 | 5.96% | 23,271 | 6.37% |
| 2 | 64,503 | 14.54% | 55,943 | 15.30% |
| 1 | 281,760 | 63.50% | 219,821 | 60.13% |

Although, as expected, the absolute number of patentees having a low number of patents has decreased, resulting in an increase of the number of patentees having a high number of patents, the relative difference compared to the total number of unique patentee names is rather moderate.

For example, the number of patentees having only one patent decreased significantly by 22% from 281,760 to 219,821 but the relative number of patentees having one patent compared to the total number of unique patentee names before and after harmonization only reduced from 63.50% to 60.13%.

Figure 2 contains the same information as Table 30 in graph format.

**Figure 2: Distribution of the number of patents per original patentee name and harmonized name**




The number of patentees has been converted to a logarithmic scale because of the range of values. Both distributions are extremely closely aligned to each other.

Although the impact of name cleaning and harmonization varies greatly at the level of individual names, the overall impact seems to be averaged out amongst all names.

Table 31 contains the share of the cumulative number of patents of top patentees for original patentee names and harmonized names.

**Table 31: Cumulative number of patents of top patentees for original patentee names and harmonized names**

| TOP PATENTEES | RELATIVE NBR PATENTS BEFORE HARMONIZATION | RELATIVE NBR PATENTS AFTER HARMONIZATION |
|---|---|---|
| Top 50 | 19% | 20% |
| Top 100 | 24% | 26% |
| Top 250 | 33% | 35% |
| Top 500 | 40% | 43% |
| Top 750 | 44% | 47% |
| Top 1000 | 47% | 49% |
| Top 5000 | 62% | 63% |
| Top 10000 | 69% | 68% |

As expected, name harmonization assigns more patents to top patentees, although the differences are rather small.

## 3.4 Patent ranking of patentees

Table 32 contains the top 25 patentees based on the original patentee names before name cleaning and harmonization. The number of patents is the sum of all EPO patent applications published between 1978 and 2004 (based on the EPO ESPACE ACCESS product) and all USPTO granted patents published between 1991 and 2003 (based on the USPTO Grant Red Book

product). Appendix 5 contains the top 200 patentees based on the original patentee names before name cleaning and harmonization.

**Table 32: Top 25 patentees before name cleaning and harmonization**

| RANK | ORIGINAL PATENTEE NAME | PAT | PAT CUM | PAT CUM PCT | PAT EPO | PAT USPTO |
|---|---|---|---|---|---|---|
| 1 | CANON KABUSHIKI KAISHA | 31649 | 31649 | 0.98% | 11293 | 20356 |
| 2 | SIEMENS AKTIENGESELLSCHAFT | 30452 | 62101 | 1.93% | 23276 | 7176 |
| 3 | INTERNATIONAL BUSINESS MACHINES CORPORATION | 28469 | 90570 | 2.82% | 2393 | 26076 |
| 4 | MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD. | 25594 | 116164 | 3.61% | 12576 | 13018 |
| 5 | SONY CORPORATION | 23620 | 139784 | 4.35% | 9358 | 14262 |
| 6 | NEC CORPORATION | 23468 | 163252 | 5.08% | 7272 | 16196 |
| 7 | KABUSHIKI KAISHA TOSHIBA | 23277 | 186529 | 5.80% | 8677 | 14600 |
| 8 | HITACHI, LTD. | 22226 | 208755 | 6.49% | 7845 | 14381 |
| 9 | GENERAL ELECTRIC COMPANY | 19117 | 227872 | 7.09% | 7762 | 11355 |
| 10 | EASTMAN KODAK COMPANY | 18847 | 246719 | 7.67% | 7672 | 11175 |
| 11 | MITSUBISHI DENKI KABUSHIKI KAISHA | 18445 | 265164 | 8.25% | 5053 | 13392 |
| 12 | FUJITSU LIMITED | 18310 | 283474 | 8.82% | 6756 | 11554 |
| 13 | ROBERT BOSCH GMBH | 16870 | 300344 | 9.34% | 11304 | 5566 |
| 14 | BASF AKTIENGESELLSCHAFT | 16855 | 317199 | 9.87% | 11883 | 4972 |
| 15 | MOTOROLA, INC. | 15758 | 332957 | 10.36% | 4043 | 11715 |
| 16 | KONINKLIJKE PHILIPS ELECTRONICS N.V. | 14411 | 347368 | 10.80% | 12374 | 2037 |
| 17 | SAMSUNG ELECTRONICS CO., LTD. | 13561 | 360929 | 11.23% | 3559 | 10002 |
| 18 | FUJI PHOTO FILM CO., LTD. | 12652 | 373581 | 11.62% | 4830 | 7822 |
| 19 | XEROX CORPORATION | 12104 | 385685 | 12.00% | 4104 | 8000 |
| 20 | INTERNATIONAL BUSINESS MACHINES<BR>CORPORATION | 11178 | 396863 | 12.34% | 11178 | |
| 21 | HEWLETT-PACKARD COMPANY | 11018 | 407881 | 12.69% | 3529 | 7489 |
| 22 | SHARP KABUSHIKI KAISHA | 10584 | 418465 | 13.02% | 4057 | 6527 |
| 23 | TEXAS INSTRUMENTS INCORPORATED | 10353 | 428818 | 13.34% | 2962 | 7391 |
| 24 | BAYER AG | 10053 | 438871 | 13.65% | 9763 | 290 |
| 25 | MINNESOTA MINING AND MANUFACTURING COMPANY | 9740 | 448611 | 13.95% | 5508 | 4232 |

Table 33 contains the top 25 patentees based on the harmonized names. Appendix 6 contains the top 200 patentees based on the harmonized names.

**Table 33: Top 25 patentees after name cleaning and harmonization**

| RANK | HARMONIZED NAME | PAT | PAT CUM | PAT CUM PCT | PAT EPO | PAT USPTO |
|---|---|---|---|---|---|---|
| 1 | INTERNATIONAL BUSINESS MACHINES CORPORATION | 41173 | 41173 | 1.28 | 13575 | 27598 |
| 2 | CANON | 31741 | 72914 | 2.27 | 11304 | 20437 |
| 3 | SIEMENS | 30770 | 103684 | 3.22 | 23398 | 7372 |
| 4 | MATSUSHITA ELECTRIC INDUSTRIAL COMPANY | 26379 | 130063 | 4.05 | 12811 | 13568 |
| 5 | SONY CORPORATION | 23665 | 153728 | 4.78 | 9358 | 14307 |
| 6 | NEC CORPORATION | 23508 | 177236 | 5.51 | 7273 | 16235 |
| 7 | TOSHIBA | 23344 | 200580 | 6.24 | 8696 | 14648 |
| 8 | HITACHI | 22754 | 223334 | 6.95 | 7934 | 14820 |
| 9 | GENERAL ELECTRIC COMPANY | 19620 | 242954 | 7.56 | 7762 | 11858 |
| 10 | EASTMAN KODAK COMPANY | 18863 | 261817 | 8.14 | 7672 | 11191 |
| 11 | FUJITSU | 18575 | 280392 | 8.72 | 6756 | 11819 |
| 12 | MITSUBISHI DENKI | 18513 | 298905 | 9.30 | 5053 | 13460 |
| 13 | BASF | 18499 | 317404 | 9.87 | 12532 | 5967 |
| 14 | MOTOROLA | 17294 | 334698 | 10.41 | 4401 | 12893 |
| 15 | BAYER | 17220 | 351918 | 10.95 | 11341 | 5879 |
| 16 | ROBERT BOSCH | 17052 | 368970 | 11.48 | 11359 | 5693 |
| 17 | SAMSUNG ELECTRONICS COMPANY | 14897 | 383867 | 11.94 | 3842 | 11055 |
| 18 | KONINKLIJKE PHILIPS ELECTRONICS | 14550 | 398417 | 12.39 | 12374 | 2176 |
| 19 | FUJI PHOTO FILM COMPANY | 12985 | 411402 | 12.80 | 4936 | 8049 |
| 20 | E.I. DU PONT DE NEMOURS & COMPANY | 12252 | 423654 | 13.18 | 6727 | 5525 |
| 21 | XEROX CORPORATION | 12111 | 435765 | 13.55 | 4104 | 8007 |
| 22 | HEWLETT-PACKARD COMPANY | 12024 | 447789 | 13.93 | 3747 | 8277 |
| 23 | THE PROCTER & GAMBLE COMPANY | 11862 | 459651 | 14.30 | 7301 | 4561 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 24 | SHARP | 10880 | 470531 | 14.64 | 4102 | 6778 |
| 25 | TEXAS INSTRUMENTS | 10801 | 481332 | 14.97 | 3247 | 7554 |

Although presenting a ranking based on the combined EPO and USPTO patents might not be very meaningful (applications versus grants, 1978-2004 versus 1991-2003), the comparison of the two rankings gives an idea of the impact of the name cleaning and harmonization procedure.

Differences in ranking are surprisingly small. "INTERNATIONAL BUSINESS MACHINES CORPORATION" jumps from 3rd and 20th place to 1st place. Places 4 to 10 remain unchanged. 11 and 12 have switched, "ROBERT BOSCH" drops from 13 to 16, "BAYER" jumps from 24 to 13, "KONINKLIJKE PHILIPS ELECTRONICS" and "SAMSUNG ELECTRONICS COMPANY" have switched, "E.I. DU PONT DE NEMOURS & COMPANY" enters at place 20, pushing "MINNESOTA MINING AND MANUFACTURING COMPANY" out of the top 25, and "THE PROCTER & GAMBLE COMPANY" enters at place 23 because of the combination of "INTERNATIONAL BUSINESS MACHINES CORPORATION and "INTERNATIONAL BUSINESS MACHINES<BR>CORPORATION".

Table 34 contains the top 25 patentees after name cleaning and harmonization with the number of matched original patentee names for the particular harmonized name, the total number of patents, the number of patents of the original patentee name having the most assigned patents, the additional number of patents assigned to the particular harmonized name, and the share of additional assigned patents in the total number of patents assigned to the particular harmonized name.

**Table 34: Additional assigned patents top 25 patentees**

| HARMONIZED NAME | NBR NAMES | NBR PAT | MAX PAT | ADD PAT | SHARE |
|---|---|---|---|---|---|
| INTERNATIONAL BUSINESS MACHINES CORPORATION | 20 | 41,173 | 28,469 | 12,704 | 31% |
| CANON | 26 | 31,741 | 31,649 | 92 | 0% |
| SIEMENS | 18 | 30,770 | 30,452 | 318 | 1% |
| MATSUSHITA ELECTRIC INDUSTRIAL COMPANY | 38 | 26,379 | 25,594 | 785 | 3% |
| SONY CORPORATION | 10 | 23,665 | 23,620 | 45 | 0% |
| NEC CORPORATION | 5 | 23,508 | 23,468 | 40 | 0% |
| TOSHIBA | 4 | 23,344 | 23,277 | 67 | 0% |
| HITACHI | 17 | 22,754 | 22,226 | 528 | 2% |
| GENERAL ELECTRIC COMPANY | 12 | 19,620 | 19,117 | 503 | 3% |
| EASTMAN KODAK COMPANY | 6 | 18,863 | 18,847 | 16 | 0% |
| FUJITSU | 10 | 18,575 | 18,310 | 265 | 1% |
| MITSUBISHI DENKI | 31 | 18,513 | 18,445 | 68 | 0% |
| BASF | 8 | 18,499 | 16,855 | 1,644 | 9% |
| MOTOROLA | 16 | 17,294 | 15,758 | 1,536 | 9% |
| BAYER | 11 | 17,220 | 10,053 | 7,167 | 42% |
| ROBERT BOSCH | 9 | 17,052 | 16,870 | 182 | 1% |
| SAMSUNG ELECTRONICS COMPANY | 50 | 14,897 | 13,561 | 1,336 | 9% |
| KONINKLIJKE PHILIPS ELECTRONICS | 11 | 14,550 | 14,411 | 139 | 1% |
| FUJI PHOTO FILM COMPANY | 27 | 12,985 | 12,652 | 333 | 3% |
| E.I. DU PONT DE NEMOURS & COMPANY | 51 | 12,252 | 6,269 | 5,983 | 49% |
| XEROX CORPORATION | 4 | 12,111 | 12,104 | 7 | 0% |
| HEWLETT-PACKARD COMPANY | 23 | 12,024 | 11,018 | 1,006 | 8% |
| THE PROCTER & GAMBLE COMPANY | 8 | 11,862 | 7,300 | 4,562 | 38% |
| SHARP | 19 | 10,880 | 10,584 | 296 | 3% |
| TEXAS INSTRUMENTS | 8 | 10,801 | 10,353 | 448 | 4% |

Table 34 gives further insight into the shift in ranking before and after name cleaning and harmonization. The jump of "INTERNATIONAL BUSINESS MACHINES CORPORATION", "BAYER", "E.I. DU PONT DE NEMOURS & COMPANY" and "THE PROCTER & GAMBLE COMPANY" can be explained by the high share of additional assigned patents from name harmonization. The relative increase in additional patents of the other top patentees is rather moderate.

The absolute differences in the number of patents are quite large for the top patentees, reducing the impact of additional patents on the relative ranking. Lower down the ranking, differences are much smaller, resulting in a much higher sensitivity to additional assigned patents because of name harmonization.

# APPENDIX 2: ALL SEARCH AND REPLACE STATEMENTS FOR ALL LEGAL FORMS TO BE REMOVED AT THE END OF A NAME

All 1.060 search and replace statements for all legal forms to be removed at the end of the name, with the keyword containing the spelling variation identified as legal form, the number of occurrences of the spelling variation in all names, the harmonized legal form, and remarks on how to replace the legal form.

| ID | KEYWORD | NBR | LEGAL FORM | REMARKS |
|---|---|---|---|---|
| 1000 | " MFG. COMPANY INC." | 2 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1001 | " MFG. COMPANY, INC." | 5 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1002 | " MFG. CO. INC." | 13 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1003 | " MFG. CO., INC." | 96 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1004 | " MFG CO., INC." | 1 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1005 | " MFG CO, INC" | 1 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1006 | " MFG. CO. INC" | 1 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1007 | " MFG CO. INC." | 1 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1008 | " MFG., CO., INC." | 1 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1009 | " MFG. CO., INC" | 1 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1010 | " MFG. CO, INC." | 1 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1011 | " MFG. CO, INC" | 1 | INCORPORATED | Replace with " MANUFACTURING COMPANY" |
| 1012 | ", CO., INC." | 37 | INCORPORATED | Replace with " COMPANY" |
| 1013 | ", CO. INC." | 4 | INCORPORATED | Replace with " COMPANY" |
| 1014 | ", CO., INC" | 1 | INCORPORATED | Replace with " COMPANY" |
| 1015 | " CO., INC." | 1497 | INCORPORATED | Replace with " COMPANY" |
| 1016 | " CO. INC." | 193 | INCORPORATED | Replace with " COMPANY" |
| 1017 | " CO., INC" | 42 | INCORPORATED | Replace with " COMPANY" |
| 1018 | " CO, INC." | 30 | INCORPORATED | Replace with " COMPANY" |
| 1019 | " CO. INC" | 5 | INCORPORATED | Replace with " COMPANY" |
| 1020 | " CO, INC" | 5 | INCORPORATED | Replace with " COMPANY" |
| 1021 | " CO.. INC." | 0 | INCORPORATED | Replace with " COMPANY" |
| 1022 | " CO INC" | 4 | INCORPORATED | Replace with " COMPANY" |
| 1023 | " CO. (INC.)" | 2 | INCORPORATED | Replace with " COMPANY" |
| 1024 | " CO INC." | 1 | INCORPORATED | Replace with " COMPANY" |
| 1025 | ", MFG., INC." | 3 | INCORPORATED | Replace with " MANUFACTURING" |
| 1026 | ", MFG. INC." | 1 | INCORPORATED | Replace with " MANUFACTURING" |
| 1027 | " MFG., INC." | 75 | INCORPORATED | Replace with " MANUFACTURING" |
| 1028 | " MFG. INC." | 40 | INCORPORATED | Replace with " MANUFACTURING" |
| 1029 | " MFG, INC." | 2 | INCORPORATED | Replace with " MANUFACTURING" |
| 1030 | " MFG., INC" | 1 | INCORPORATED | Replace with " MANUFACTURING" |
| 1031 | " MFG INC." | 2 | INCORPORATED | Replace with " MANUFACTURING" |
| 1032 | ", LTD., INC." | 9 | INCORPORATED | Remove |
| 1033 | ", LTD. INC." | 3 | INCORPORATED | Remove |
| 1034 | " LTD., INC." | 30 | INCORPORATED | Remove |
| 1035 | " LTD. INC." | 4 | INCORPORATED | Remove |
| 1036 | " LTD, INC." | 3 | INCORPORATED | Remove |
| 1037 | " LTD INC." | 2 | INCORPORATED | Remove |
| 1038 | ", INTL., INC." | 3 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1039 | ", INT'L., INC." | 2 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1040 | ", INT'L. INC." | 1 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1041 | " INTL., INC." | 3 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1042 | " INT'L., INC." | 8 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1043 | " INT'L. INC." | 1 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1044 | " INT'L, INC." | 14 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1045 | " INT'L INC." | 10 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1046 | " INTL, INC." | 2 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1047 | " INTL. INC." | 2 | INCORPORATED | Replace with " INTERNATIONAL" |
| 1048 | " CORP. INC." | 7 | INCORPORATED | Replace with " CORPORATION" |
| 1049 | " CORP., INC." | 16 | INCORPORATED | Replace with " CORPORATION" |
| 1050 | " CORP., INC" | 1 | INCORPORATED | Replace with " CORPORATION" |
| 1051 | ", INC." | 55347 | INCORPORATED | Remove |
| 1052 | ", INC" | 1536 | INCORPORATED | Remove |
| 1053 | ", INC.." | 1 | INCORPORATED | Remove |
| 1054 | "; INC." | 6 | INCORPORATED | Remove |
| 1055 | ", INC/" | 1 | INCORPORATED | Remove |

| | | | | |
|---|---|---|---|---|
| 1056 | ", IN.C" | 4 | INCORPORATED | Remove |
| 1057 | " INC." | 14982 | INCORPORATED | Remove |
| 1058 | " INC" | 835 | INCORPORATED | Remove |
| 1059 | ", MFG. CO., LTD." | 4 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1060 | ", MFG., CO., LTD." | 1 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1061 | " (MFG) CO., LTD." | 1 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1062 | " (MFG.) CO., LTD." | 1 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1063 | " MFG. CO., LTD." | 221 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1064 | " MFG. CO., LTD" | 37 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1065 | " MFG CO., LTD." | 12 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1066 | " MFG CO., LTD" | 5 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1067 | " MFG. CO. LTD." | 70 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1068 | " MFG., CO., LTD." | 32 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1069 | " MFG., CO. LTD." | 8 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1070 | " MFG, CO., LTD." | 7 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1071 | " MFG CO. LTD." | 5 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1072 | " MFG. CO. LTD" | 7 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1073 | " MFG CO. LTD" | 1 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1074 | " MFG., CO., LTD" | 2 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1075 | " MFG. CO, LTD." | 4 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1076 | " MFG CO LTD" | 3 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1077 | " MFG CO LTD." | 1 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1078 | " MFG. CO. LTD.." | 0 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1079 | " M.F.G. CO., LTD." | 1 | LIMITED | Replace with " MANUFACTURING COMPANY" |
| 1080 | " INT. CO., LTD." | 2 | LIMITED | Replace with " INTERNATIONAL COMPANY" |
| 1081 | " INT. CO. LTD." | 1 | LIMITED | Replace with " INTERNATIONAL COMPANY" |
| 1082 | " INT., CO., LTD." | 1 | LIMITED | Replace with " INTERNATIONAL COMPANY" |
| 1083 | " INT'L CO., LTD." | 4 | LIMITED | Replace with " INTERNATIONAL COMPANY" |
| 1084 | " CO. CO., LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1085 | " CO CO., LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1086 | " CO., CO. LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1087 | " CO., CO., LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1088 | ", CO., LTD." | 409 | LIMITED | Replace with " COMPANY" |
| 1089 | ", CO. LTD." | 58 | LIMITED | Replace with " COMPANY" |
| 1090 | ", CO., LTD" | 29 | LIMITED | Replace with " COMPANY" |
| 1091 | ", CO, LTD." | 10 | LIMITED | Replace with " COMPANY" |
| 1092 | ", CO. LTD" | 9 | LIMITED | Replace with " COMPANY" |
| 1093 | ", CO LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1094 | ", CO, LTD" | 1 | LIMITED | Replace with " COMPANY" |
| 1095 | ", CO.. LTD." | 0 | LIMITED | Replace with " COMPANY" |
| 1096 | " CO., LTD." | 11390 | LIMITED | Replace with " COMPANY" |
| 1097 | " CO. LTD." | 1888 | LIMITED | Replace with " COMPANY" |
| 1098 | " CO., LTD" | 1197 | LIMITED | Replace with " COMPANY" |
| 1099 | " CO, LTD." | 185 | LIMITED | Replace with " COMPANY" |
| 1100 | " CO. LTD" | 217 | LIMITED | Replace with " COMPANY" |
| 1101 | " CO LTD." | 33 | LIMITED | Replace with " COMPANY" |
| 1102 | " CO, LTD" | 42 | LIMITED | Replace with " COMPANY" |
| 1103 | " CO LTD" | 111 | LIMITED | Replace with " COMPANY" |
| 1104 | " CO., LT.D." | 1 | LIMITED | Replace with " COMPANY" |
| 1105 | " CO.. LTD." | 0 | LIMITED | Replace with " COMPANY" |
| 1106 | " CO. L.T.D." | 1 | LIMITED | Replace with " COMPANY" |
| 1107 | " CO; LTD." | 2 | LIMITED | Replace with " COMPANY" |
| 1108 | " CO., L.T.D." | 3 | LIMITED | Replace with " COMPANY" |
| 1109 | " CO,, LTD." | 4 | LIMITED | Replace with " COMPANY" |
| 1110 | " CO.., LTD." | 0 | LIMITED | Replace with " COMPANY" |
| 1111 | " CO.?, LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1112 | " CO; LTD" | 1 | LIMITED | Replace with " COMPANY" |
| 1113 | " CO:, LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1114 | " CO., LTD.." | 0 | LIMITED | Replace with " COMPANY" |
| 1115 | " CO.; LTD." | 2 | LIMITED | Replace with " COMPANY" |
| 1116 | " CO,, LTD" | 1 | LIMITED | Replace with " COMPANY" |
| 1117 | " CO.; LTD" | 1 | LIMITED | Replace with " COMPANY" |
| 1118 | " CO., LTD,." | 0 | LIMITED | Replace with " COMPANY" |
| 1119 | " CO., LT.D" | 1 | LIMITED | Replace with " COMPANY" |
| 1120 | " CO.., LTD" | 1 | LIMITED | Replace with " COMPANY" |
| 1121 | " CO,., LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1122 | " CO: LTD" | 1 | LIMITED | Replace with " COMPANY" |
| 1123 | " CO. PTY. LTD." | 27 | LIMITED | Replace with " COMPANY" |
| 1124 | " CO. PTY LTD" | 2 | LIMITED | Replace with " COMPANY" |
| 1125 | " CO. PTY LTD." | 4 | LIMITED | Replace with " COMPANY" |
| 1126 | " CO PTY LTD" | 5 | LIMITED | Replace with " COMPANY" |

| | | | | |
|---|---|---|---|---|
| 1127 | " CO PTY LTD." | 2 | LIMITED | Replace with " COMPANY" |
| 1128 | " CO., PTY. LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1129 | " CO., PTY., LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1130 | " CO., PTY LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1131 | " (INT'L) PTY. LTD." | 1 | LIMITED | Replace with " (INTERNATIONAL)" |
| 1132 | " (INTL.) PTY. LTD." | 1 | LIMITED | Replace with " (INTERNATIONAL)" |
| 1133 | " MFG. PTY. LTD." | 1 | LIMITED | Replace with " MANUFACTURING" |
| 1134 | " MFG. PTY. LTD" | 1 | LIMITED | Replace with " MANUFACTURING" |
| 1135 | " (QLD) PTY. LTD" | 1 | LIMITED | Remove |
| 1136 | " (QLD.) PTY. LTD." | 2 | LIMITED | Remove |
| 1137 | " QLD PTY LTD." | 1 | LIMITED | Remove |
| 1138 | " (QLD) PTY LTD" | 2 | LIMITED | Remove |
| 1139 | " (QLD) PTY. LTD." | 2 | LIMITED | Remove |
| 1140 | " (QLD) PTY LTD." | 1 | LIMITED | Remove |
| 1141 | " (VIC) PTY., LTD." | 1 | LIMITED | Remove |
| 1142 | " (VIC) PTY LTD." | 2 | LIMITED | Remove |
| 1143 | " (VIC) PTY. LTD." | 2 | LIMITED | Remove |
| 1144 | " (VIC) PTY LTD" | 1 | LIMITED | Remove |
| 1145 | " (VIC.) PTY. LTD." | 2 | LIMITED | Remove |
| 1146 | " (S.A.) PTY LTD" | 1 | LIMITED | Remove |
| 1147 | " (SA) PTY LTD" | 2 | LIMITED | Remove |
| 1148 | " S.A. (PTY) LTD." | 2 | LIMITED | Remove |
| 1149 | ", PTY. LTD." | 9 | LIMITED | Remove |
| 1150 | ", PTY LTD" | 4 | LIMITED | Remove |
| 1151 | ", PTY., LTD." | 2 | LIMITED | Remove |
| 1152 | ", PTY LTD." | 2 | LIMITED | Remove |
| 1153 | ", PTY, LTD." | 4 | LIMITED | Remove |
| 1154 | ", PTY, LTD" | 1 | LIMITED | Remove |
| 1155 | " PTY. LTD." | 1425 | LIMITED | Remove |
| 1156 | " PTY LTD" | 861 | LIMITED | Remove |
| 1157 | " PTY., LTD." | 79 | LIMITED | Remove |
| 1158 | " PTY LTD." | 672 | LIMITED | Remove |
| 1159 | " PTY, LTD." | 73 | LIMITED | Remove |
| 1160 | " PTY, LTD" | 20 | LIMITED | Remove |
| 1161 | " PTY. LTD" | 78 | LIMITED | Remove |
| 1162 | " (PTY) LTD" | 36 | LIMITED | Remove |
| 1163 | " (PTY) LTD." | 48 | LIMITED | Remove |
| 1164 | " PTY., LTD" | 4 | LIMITED | Remove |
| 1165 | " (PTY.) LTD." | 2 | LIMITED | Remove |
| 1166 | " PTY: LTD." | 1 | LIMITED | Remove |
| 1167 | " (PTY.) LTD" | 1 | LIMITED | Remove |
| 1168 | " (PTY), LTD." | 1 | LIMITED | Remove |
| 1169 | " CO. PTE. LTD." | 2 | LIMITED | Replace with " COMPANY" |
| 1170 | " CO. (PTE) LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1171 | " CO. PTE LTD" | 2 | LIMITED | Replace with " COMPANY" |
| 1172 | " CO., PTE. LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1173 | " (S) PTE LTD." | 7 | LIMITED | Remove |
| 1174 | " (S) PTE LTD" | 6 | LIMITED | Remove |
| 1175 | " (S) PTE. LTD." | 3 | LIMITED | Remove |
| 1176 | " (S), PTE., LTD." | 1 | LIMITED | Remove |
| 1177 | ", PTE., LTD." | 1 | LIMITED | Remove |
| 1178 | ", PTE. LTD." | 3 | LIMITED | Remove |
| 1179 | ", PTE LTD" | 1 | LIMITED | Remove |
| 1180 | ", PTE LTD." | 5 | LIMITED | Remove |
| 1181 | ", PTE, LTD." | 5 | LIMITED | Remove |
| 1182 | " PTE., LTD." | 5 | LIMITED | Remove |
| 1183 | " PTE. LTD." | 87 | LIMITED | Remove |
| 1184 | " PTE LTD" | 78 | LIMITED | Remove |
| 1185 | " PTE LTD." | 112 | LIMITED | Remove |
| 1186 | " PTE, LTD." | 21 | LIMITED | Remove |
| 1187 | " PTE. LTD" | 9 | LIMITED | Remove |
| 1188 | " (PTE) LTD." | 9 | LIMITED | Remove |
| 1189 | " (PTE) LTD" | 2 | LIMITED | Remove |
| 1190 | " PTE, LTD" | 2 | LIMITED | Remove |
| 1191 | ", CORP. LTD." | 1 | LIMITED | Replace with " CORPORATION" |
| 1192 | " CORP. LTD." | 23 | LIMITED | Replace with " CORPORATION" |
| 1193 | " CORP., LTD." | 27 | LIMITED | Replace with " CORPORATION" |
| 1194 | " CORP, LTD." | 3 | LIMITED | Replace with " CORPORATION" |
| 1195 | " CORP., LTD" | 2 | LIMITED | Replace with " CORPORATION" |
| 1196 | " CORP. LTD" | 1 | LIMITED | Replace with " CORPORATION" |
| 1197 | " MFG LTD." | 1 | LIMITED | Replace with " MANUFACTURING" |

| | | | | |
|---|---|---|---|---|
| 1199 | " MFG., LTD." | 23 | LIMITED | Replace with " MANUFACTURING" |
| 1200 | " MFG. LTD." | 13 | LIMITED | Replace with " MANUFACTURING" |
| 1201 | " MFG., LTD" | 1 | LIMITED | Replace with " MANUFACTURING" |
| 1202 | " CO., INC. LTD." | 2 | LIMITED | Replace with " COMPANY" |
| 1203 | " CO., INC., LTD." | 1 | LIMITED | Replace with " COMPANY" |
| 1204 | ", INC., LTD" | 1 | LIMITED | Remove |
| 1205 | ", INC., LTD." | 4 | LIMITED | Remove |
| 1206 | ", INC. LTD." | 2 | LIMITED | Remove |
| 1207 | " INC., LTD" | 0 | LTD | Remove |
| 1208 | " INC., LTD." | 3 | LIMITED | Remove |
| 1209 | " INC. LTD." | 4 | LIMITED | Remove |
| 1210 | " INC. LTD" | 2 | LIMITED | Remove |
| 1211 | " INC, LTD." | 1 | LIMITED | Remove |
| 1212 | " INT'L LTD." | 1 | LIMITED | Replace with " INTERNATIONAL" |
| 1213 | " INT'L. LTD." | 1 | LIMITED | Replace with " INTERNATIONAL" |
| 1214 | " INT'L., LTD." | 1 | LIMITED | Replace with " INTERNATIONAL" |
| 1215 | " INT""L LTD." | 1 | LIMITED | Replace with " INTERNATIONAL" |
| 1216 | " INTL. LTD." | 2 | LIMITED | Replace with " INTERNATIONAL" |
| 1217 | " INT., LTD." | 2 | LIMITED | Replace with " INTERNATIONAL" |
| 1218 | " INT. LTD." | 2 | LIMITED | Replace with " INTERNATIONAL" |
| 1219 | " KABUSHIKI KAISHA, LTD." | 3 | LIMITED | Remove |
| 1220 | ", LTD." | 4316 | LIMITED | Remove |
| 1221 | ", LTD" | 324 | LIMITED | Remove |
| 1222 | ", LTD.." | 0 | LIMITED | Remove |
| 1223 | ", L.T.D." | 3 | LIMITED | Remove |
| 1224 | " (LTD.)" | 7 | LIMITED | Remove |
| 1225 | " LTD." | 8778 | LIMITED | Remove |
| 1226 | " LTD" | 1620 | LIMITED | Remove |
| 1227 | " LTD.." | 1 | LIMITED | Remove |
| 1228 | " L.T.D." | 8 | LIMITED | Remove |
| 1229 | " (LTD)" | 2 | LIMITED | Remove |
| 1230 | ", LTD/" | 1 | LIMITED | Remove |
| 1231 | " & C. S.P.A." | 51 | SPA | Replace with " & COMPANY" only if "C" is preceeded by "&" |
| 1232 | " & C. SPA" | 8 | SPA | Replace with " & COMPANY" only if "C" is preceeded by "&" |
| 1233 | " & C SPA" | 1 | SPA | Replace with " & COMPANY" only if "C" is preceeded by "&" |
| 1234 | " & C., S.P.A." | 6 | SPA | Replace with " & COMPANY" only if "C" is preceeded by "&" |
| 1235 | " & C. S.P.A" | 1 | SPA | Replace with " & COMPANY" only if "C" is preceeded by "&" |
| 1236 | " & C. -S.P.A." | 1 | SPA | Replace with " & COMPANY" only if "C" is preceeded by "&" |
| 1237 | " CO. S.P.A." | 4 | SPA | Replace with " COMPANY" |
| 1238 | " CO. SPA" | 1 | SPA | Replace with " COMPANY" |
| 1239 | " CO., S.P.A." | 1 | SPA | Replace with " COMPANY" |
| 1240 | ", S.P.A." | 227 | SPA | Remove |
| 1241 | ", SPA" | 23 | SPA | Remove |
| 1242 | ", S.P.A" | 6 | SPA | Remove |
| 1243 | " - S.P.A" | 3 | SPA | Remove |
| 1244 | " -S.P.A." | 3 | SPA | Remove |
| 1245 | ", SPA." | 1 | SPA | Remove |
| 1246 | " S.P.A." | 4819 | SPA | Remove |
| 1247 | " SPA" | 598 | SPA | Remove |
| 1248 | " S.P.A" | 188 | SPA | Remove |
| 1249 | " SPA." | 10 | SPA | Remove |
| 1250 | " S.P.A.." | 2 | SPA | Remove |
| 1251 | " S.PA." | 1 | SPA | Remove |
| 1252 | " S-P.A." | 1 | SPA | Remove |
| 1253 | " (S.P.A.)" | 1 | SPA | Remove |
| 1254 | " SP.A." | 1 | SPA | Remove |
| 1255 | " SPA'" | 1 | SPA | Remove |
| 1256 | " & C. S.R.L." | 34 | SRL | Replace with " & COMPANY" only if "C" is preceeded by "&" or "E" |
| 1257 | " & C. SRL" | 5 | SRL | Replace with " & COMPANY" only if "C" is preceeded by "&" or "E" |
| 1258 | " & C. S.R.L" | 1 | SRL | Replace with " & COMPANY" only if "C" is preceeded by "&" or "E" |
| 1259 | " E C. S.R.L." | 3 | SRL | Replace with " & COMPANY" only if "C" is preceeded by "&" or "E" |

| | | | | |
|---|---|---|---|---|
| 1260 | " & C S.R.L." | 1 | SRL | Replace with " & COMPANY" only if "C" is preceeded by "&" or "E" |
| 1261 | " CO. S.R.L." | 5 | SRL | Replace with " COMPANY" |
| 1262 | " LTD. SRL" | 1 | SRL | Remove |
| 1263 | " L.T.D. S.R.L." | 1 | SRL | Remove |
| 1264 | ", S.R.L." | 103 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1265 | " - S.R.L." | 53 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1266 | ", S.R.L" | 4 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1267 | ", SRL" | 8 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1268 | ", SRL." | 3 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1269 | ", S.RL" | 1 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1270 | " S.R.L." | 4659 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1271 | " S.R.L" | 109 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1272 | " SRL" | 470 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1273 | " SRL." | 25 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1274 | " S.RL" | 0 | SRL | Remove |
| 1275 | " -S.R.L." | 4 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1276 | " .S.R.L." | 0 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1277 | " SR.L." | 1 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1278 | " S.RL." | 2 | SRL | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1279 | " S.ß.R.L." | 3 | S.ß.R.L. | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1280 | " S.ß.R.L" | 1 | S.ß.R.L. | Remove and label as SRL unless "S.ß.R.L", label as S.ß.R.L |
| 1281 | " CO. PTY. LIMITED" | 7 | LIMITED | Replace with " COMPANY" |
| 1282 | " CO. PTY LIMITED" | 1 | LIMITED | Replace with " COMPANY" |
| 1283 | " CO PTY LIMITED" | 3 | LIMITED | Replace with " COMPANY" |
| 1284 | " CO. (PTY) LIMITED" | 1 | LIMITED | Replace with " COMPANY" |
| 1285 | ", PTY. LIMITED" | 1 | LIMITED | Remove |
| 1286 | ", PTY, LIMITED" | 1 | LIMITED | Remove |
| 1287 | " PTY. LIMITED" | 479 | LIMITED | Remove |
| 1288 | " PTY, LIMITED" | 9 | LIMITED | Remove |
| 1289 | " PTY LIMITED" | 417 | LIMITED | Remove |
| 1290 | " (PTY) LIMITED" | 72 | LIMITED | Remove |
| 1291 | " PTY., LIMITED" | 4 | LIMITED | Remove |
| 1292 | " PTY. LIMITED." | 2 | LIMITED | Remove |
| 1293 | " PTY. .LIMITED" | 0 | LIMITED | Remove |
| 1294 | " (PTY.) LIMITED" | 2 | LIMITED | Remove |
| 1295 | " CO., LIMITED" | 82 | LIMITED | Replace with " COMPANY" |
| 1296 | " CO. LIMITED" | 129 | LIMITED | Replace with " COMPANY" |
| 1297 | " CO. LIMITED." | 1 | LIMITED | Replace with " COMPANY" |
| 1298 | " CO LIMITED" | 18 | LIMITED | Replace with " COMPANY" |
| 1299 | " CO, LIMITED" | 4 | LIMITED | Replace with " COMPANY" |
| 1300 | " CO., LIMITED." | 1 | LIMITED | Replace with " COMPANY" |
| 1301 | " CO. (NZ) LIMITED" | 1 | LIMITED | Replace with " COMPANY" |
| 1302 | " (NZ) LIMITED" | 19 | LIMITED | Remove |
| 1303 | " (N.Z.) LIMITED" | 3 | LIMITED | Remove |
| 1304 | " NZ LIMITED" | 9 | LIMITED | Remove |
| 1305 | " (H.K.) LIMITED" | 11 | LIMITED | Remove |
| 1306 | " (H.K) LIMITED" | 1 | LIMITED | Remove |
| 1307 | " (HK) LIMITED" | 6 | LIMITED | Remove |
| 1308 | " HK LIMITED" | 2 | LIMITED | Remove |
| 1309 | " (IP) LIMITED" | 4 | LIMITED | Remove |
| 1310 | " (I.P.) LIMITED" | 3 | LIMITED | Remove |
| 1311 | " I.P. LIMITED" | 2 | LIMITED | Remove |
| 1312 | " IP LIMITED" | 9 | LIMITED | Remove |
| 1313 | " ( IP) LIMITED" | 1 | LIMITED | Remove |

| | | | | |
|---|---|---|---|---|
| 1314 | " (I.P) LIMITED" | 1 | LIMITED | Remove |
| 1315 | " (PTE) LIMITED" | 2 | LIMITED | Remove |
| 1316 | " PTE LIMITED" | 12 | LIMITED | Remove |
| 1317 | " PTE. LIMITED" | 3 | LIMITED | Remove |
| 1318 | " PTE, LIMITED" | 1 | LIMITED | Remove |
| 1319 | " (BVI) LIMITED" | 10 | LIMITED | Remove |
| 1320 | " (B.V.I.) LIMITED" | 2 | LIMITED | Remove |
| 1321 | " (BVIØ) LIMITED" | 1 | LIMITED | Remove |
| 1322 | " (N.I.) LIMITED" | 6 | LIMITED | Remove |
| 1323 | " (NI) LIMITED" | 1 | LIMITED | Remove |
| 1324 | " NI LIMITED" | 1 | LIMITED | Remove |
| 1325 | ", LIMITED." | 8 | LIMITED | Remove |
| 1326 | ", LIMITED" | 625 | LIMITED | Remove |
| 1327 | " LIMITED." | 21 | LIMITED | Remove |
| 1328 | " LIMITED" | 18429 | LIMITED | Remove |
| 1329 | " (PROPRIETARY LIMITED)" | 5 | LIMITED | Remove |
| 1330 | " (PROPRIETARY) (LIMITED)" | 1 | LIMITED | Remove |
| 1331 | " CIE, S. A." | 1 | SA | Replace with " COMPAGNIE" |
| 1332 | " CIE S. A." | 1 | SA | Replace with " COMPAGNIE" |
| 1333 | " FRANCE S. A." | 6 | SA | Replace with " FRANCE" |
| 1334 | " (FRANCE) S. A." | 1 | SA | Replace with " (FRANCE)" |
| 1335 | " S. P. A." | 18 | SPA | Remove, almost no interference with private person name abbreviations because already 3 characters |
| 1336 | " S P A" | 1 | SPA | Remove, almost no interference with private person name abbreviations because already 3 characters |
| 1337 | " U. CO. GMBH" | 3 | GMBH | Replace with " & COMPANY" |
| 1338 | " UND CO. GMBH" | 1 | GMBH | Replace with " & COMPANY" |
| 1339 | " CO. GMBH" | 139 | GMBH | Replace with " COMPANY" |
| 1340 | " + CO., GMBH" | 1 | GMBH | Replace with " & COMPANY" |
| 1341 | " + CO GMBH" | 1 | GMBH | Replace with " & COMPANY" |
| 1342 | " CO., GMBH" | 11 | GMBH | Replace with " COMPANY" |
| 1343 | " CO. (GMBH)" | 1 | GMBH | Replace with " COMPANY" |
| 1344 | " CO., (GMBH)" | 1 | GMBH | Replace with " COMPANY" |
| 1345 | " CO GMBH" | 10 | GMBH | Replace with " COMPANY" |
| 1346 | " &CO. GMBH" | 1 | GMBH | Replace with " & COMPANY" |
| 1347 | " CO, GMBH" | 1 | GMBH | Replace with " COMPANY" |
| 1348 | " CIE. GMBH" | 9 | GMBH | Replace with " COMPANY" |
| 1349 | " CIE, GMBH" | 1 | GMBH | Replace with " COMPANY" |
| 1350 | " CIE GMBH" | 5 | GMBH | Replace with " COMPANY" |
| 1351 | ", GMBH" | 165 | GMBH | Remove |
| 1352 | ", GMBH." | 2 | GMBH | Remove |
| 1353 | " GMBH" | 16974 | GMBH | Remove |
| 1354 | " G.M.B.H." | 75 | GMBH | Remove |
| 1355 | " GMBH." | 65 | GMBH | Remove |
| 1356 | " -GMBH" | 1 | GMBH | Remove |
| 1357 | " GESELLSCHAFT MIT BESCHRAENKTER HAFTUNG (GMBH)" | 3 | GMBH | Remove |
| 1358 | " (GMBH)" | 2 | GMBH | Remove |
| 1359 | " G.M.B.H" | 8 | GMBH | Remove |
| 1360 | " G.MBH" | 1 | GMBH | Remove |
| 1361 | " G.M.BH" | 1 | GMBH | Remove |
| 1362 | " MFG. CORPORATION" | 9 | | Replace with " MANUFACTURING CORPORATION" |
| 1363 | " ET CIE. S.A." | 2 | SA | Replace with " & COMPAGNIE" |
| 1364 | " ET CIE, S.A." | 2 | SA | Replace with " & COMPAGNIE" |
| 1365 | " ET CIE (SA)" | 1 | SA | Replace with " & COMPAGNIE" |
| 1366 | " ET CIE S.A." | 5 | SA | Replace with " & COMPAGNIE" |
| 1367 | " ET. CIE S.A." | 1 | SA | Replace with " & COMPAGNIE" |
| 1368 | " ET CIE., S.A." | 1 | SA | Replace with " & COMPAGNIE" |
| 1369 | " CIE. S.A." | 5 | SA | Replace with " COMPAGNIE" |
| 1370 | " CIE (S.A.)" | 2 | SA | Replace with " COMPAGNIE" |
| 1371 | " CIE S.A." | 14 | SA | Replace with " COMPAGNIE" |
| 1372 | " CIE, S.A." | 6 | SA | Replace with " COMPAGNIE" |
| 1373 | " CIE, SA" | 1 | SA | Replace with " COMPAGNIE" |
| 1374 | " CI.E. SA" | 1 | SA | Replace with " COMPAGNIE" |
| 1375 | " CIE SA" | 9 | SA | Replace with " COMPAGNIE" |
| 1376 | " ET CO. S.A." | 2 | SA | Replace with " & COMPANY" |
| 1377 | " ET CO S.A." | 2 | SA | Replace with " & COMPANY" |
| 1378 | " ET CO. SA" | 1 | SA | Replace with " & COMPANY" |

| | | | | |
|---|---|---|---|---|
| 1379 | " CO., S.A." | 6 | SA | Replace with " COMPANY" |
| 1382 | " CO. S.A." | 6 | SA | Replace with " COMPANY" |
| 1383 | " CO SA" | 1 | SA | Replace with " COMPANY" |
| 1384 | " CO. SA" | 1 | SA | Replace with " COMPANY" |
| 1385 | " Y CIA., S.A." | 3 | SA | Replace with " & COMPANIA" |
| 1386 | " Y CIA. S.A." | 1 | SA | Replace with " & COMPANIA" |
| 1387 | " Y CIA S.A." | 1 | SA | Replace with " & COMPANIA" |
| 1388 | " Y CIA, S.A." | 1 | SA | Replace with " & COMPANIA" |
| 1389 | " CIA, S.A." | 1 | SA | Replace with " COMPANIA" |
| 1390 | " CIA. S.A." | 3 | SA | Replace with " COMPANIA" |
| 1391 | " CIA., S.A." | 2 | SA | Replace with " COMPANIA" |
| 1392 | " CIA S.A." | 2 | SA | Replace with " COMPANIA" |
| 1393 | ", INC. S.A." | 1 | SA | Remove |
| 1394 | ", INC., SA." | 1 | SA | Remove |
| 1395 | " INC. S.A." | 1 | SA | Remove |
| 1396 | " INC., SA." | 1 | SA | Remove |
| 1397 | " INC., S.A." | 1 | SA | Remove |
| 1398 | " MFG. CY, S.A." | 1 | SA | Replace with " MANUFACTURING COMPANY" |
| 1399 | " CY, S.A." | 2 | SA | Replace with "COMPANY" |
| 1400 | " MANUFACTURIN CY, S.A." | 0 | SA | Replace with " MANUFACTURING COMPANY" |
| 1401 | " CY S.A." | 1 | SA | Replace with "COMPANY" |
| 1402 | ", S.A." | 1735 | SA | Remove |
| 1403 | ", SA" | 86 | SA | Remove |
| 1404 | ", S.A" | 11 | SA | Remove |
| 1405 | ", S,A." | 1 | SA | Remove |
| 1406 | ", SA." | 3 | SA | Remove |
| 1407 | " S.A." | 5255 | SA | Remove |
| 1408 | " SA" | 1507 | SA | Remove |
| 1409 | " S.A" | 60 | SA | Remove |
| 1410 | " S,A." | 1 | SA | Remove |
| 1411 | " SA." | 24 | SA | Remove |
| 1412 | " (S.A.)" | 142 | SA | Remove |
| 1413 | " (SA)" | 36 | SA | Remove |
| 1414 | " (S.A)" | 2 | SA | Remove |
| 1415 | " S..A." | 1 | SA | Remove |
| 1416 | " S.A.." | 1 | SA | Remove |
| 1417 | " -SA" | 1 | SA | Remove |
| 1418 | " S,A" | 1 | SA | Remove |
| 1419 | " -S.A." | 1 | SA | Remove |
| 1420 | " .S.A" | 0 | SA | Remove |
| 1421 | " S-A." | 1 | SA | Remove |
| 1422 | " S/A" | 51 | SA | Remove |
| 1423 | " S/A." | 4 | SA | Remove |
| 1424 | " E C." | 57 | | Replace with " & COMPANY" only if string ends with " E C." |
| 1425 | " P. L. C." | 2 | PLC | Remove |
| 1426 | " P L C" | 2 | PLC | Remove |
| 1427 | " P. L. C" | 1 | PLC | Remove |
| 1428 | ", L. L. C." | 4 | LLC | Remove |
| 1429 | ", L L C" | 1 | LLC | Remove |
| 1430 | " L L C." | 1 | LLC | Remove |
| 1431 | " L L C" | 2 | LLC | Remove |
| 1432 | " & C." | 426 | | Replace with " COMPANY" only if string ends with " & C." or " & C" |
| 1434 | " & C" | 18 | | Replace with " COMPANY" only if string ends with " & C." or " & C" |
| 1435 | " S. R. L." | 19 | SRL | Remove |
| 1436 | " S.R L." | 1 | SRL | Remove |
| 1437 | " S.R. L." | 5 | SRL | Remove |
| 1438 | " GES. M. B. H." | 2 | GMBH | Remove |
| 1439 | " GESELLSCHAFT M. B. H." | 4 | GMBH | Remove |
| 1440 | " CO. PLC" | 3 | PLC | Replace with " COMPANY" |
| 1441 | " CO., PLC" | 1 | PLC | Replace with " COMPANY" |
| 1442 | " CO., P.L.C." | 1 | PLC | Replace with " COMPANY" |
| 1443 | " CO. P.L.C." | 3 | PLC | Replace with " COMPANY" |
| 1444 | ", PLC" | 87 | PLC | Remove |
| 1445 | ", PLC." | 14 | PLC | Remove |
| 1446 | ", P.L.C." | 11 | PLC | Remove |
| 1447 | ", P.L.C" | 2 | PLC | Remove |
| 1448 | " PLC" | 794 | PLC | Remove |
| 1449 | " PLC." | 95 | PLC | Remove |

| | | | | |
|---|---|---|---|---|
| 1450 | " P.L.C." | 60 | PLC | Remove |
| 1451 | " P.L.C" | 2 | PLC | Remove |
| 1452 | " PL.C" | 1 | PLC | Remove |
| 1453 | " P.L.C.." | 0 | PLC | Remove |
| 1454 | " (PLC)" | 1 | PLC | Remove |
| 1455 | " ET CIE (SARL)" | 1 | SARL | Replace with " & COMPAGNIE" |
| 1456 | " CIE, SARL" | 2 | SARL | Replace with " COMPAGNIE" |
| 1457 | " CIE S.A.R.L." | 2 | SARL | Replace with " COMPAGNIE" |
| 1458 | " CIE SARL" | 3 | SARL | Replace with " COMPAGNIE" |
| 1459 | ", SARL" | 68 | SARL | Remove |
| 1460 | ", S.A.R.L." | 41 | SARL | Remove |
| 1461 | ", (SARL)" | 1 | SARL | Remove |
| 1462 | ", S.A.R.L" | 1 | SARL | Remove |
| 1463 | " SARL" | 342 | SARL | Remove |
| 1464 | " S.A.R.L." | 428 | SARL | Remove |
| 1465 | " (SARL)" | 60 | SARL | Remove |
| 1467 | " S.A.R.L" | 10 | SARL | Remove |
| 1468 | " (S.A.R.L.)" | 73 | SARL | Remove |
| 1469 | " (S.A.R.L)" | 1 | SARL | Remove |
| 1470 | " SARL." | 6 | SARL | Remove |
| 1471 | " S.A.R:L" | 1 | SARL | Remove |
| 1472 | " S.AR.L." | 1 | SARL | Remove |
| 1473 | " -SARL" | 1 | SARL | Remove |
| 1474 | " ( SARL)" | 1 | SARL | Remove |
| 1475 | " (SARL.)" | 1 | SARL | Remove |
| 1476 | " + CO<br>AKTIENGESELLSCHAFT" | 1 | AG | Replace with " & COMPANY" |
| 1477 | " + CO.<br>AKTIENGESELLSCHAFT" | 1 | AG | Replace with " & COMPANY" |
| 1478 | " CO.<br>AKTIENGESELLSCHAFT" | 9 | AG | Replace with " COMPANY" |
| 1479 | " CO AKTIENGESELLSCHAFT" | 3 | AG | Replace with " COMPANY" |
| 1480 | " CO.,<br>AKTIENGESELLSCHAFT" | 2 | AG | Replace with " COMPANY" |
| 1481 | " A.G.<br>AKTIENGESELLSCHAFT" | 1 | AG | Remove |
| 1482 | " AG AKTIENGESELLSCHAFT" | 3 | AG | Remove |
| 1483 | " CIE.<br>AKTIENGESELLSCHAFT" | 1 | AG | Replace with " COMPAGNIE" |
| 1484 | " CIE<br>AKTIENGESELLSCHAFT" | 2 | AG | Replace with " COMPAGNIE" |
| 1485 | ", AKTIENGESELLSCHAFT" | 6 | AG | Remove |
| 1486 | " AKTIENGESELLSCHAFT" | 1141 | AG | Remove |
| 1487 | " AKTIENGESELL-SCHAFT" | 1 | AG | Remove |
| 1488 | " AKTIEN-GESELLSCHAFT" | 3 | AG | Remove |
| 1489 | " EN CO. N.V." | 1 | NV | Replace with " & COMPANY" |
| 1490 | " EN CO. NV" | 1 | NV | Replace with " & COMPANY" |
| 1491 | " CO. N.V." | 3 | NV | Replace with " COMPANY" |
| 1492 | " CO N.V." | 1 | NV | Replace with " COMPANY" |
| 1493 | " N.A. N.V." | 2 | NV | Remove |
| 1494 | " (NA) N.V." | 1 | NV | Remove |
| 1495 | " (NA) NV" | 2 | NV | Remove |
| 1496 | " (N.A.) N.V." | 1 | NV | Remove |
| 1497 | ", INC. N.V." | 1 | NV | Remove |
| 1498 | ", INC. (NV)" | 1 | NV | Remove |
| 1499 | " INC. NV." | 1 | NV | Remove |
| 1500 | " INC. N.V." | 1 | NV | Remove |
| 1501 | " CORP. N.V." | 3 | NV | Replace with " CORPORATION" |
| 1502 | ", N.V." | 129 | NV | Remove |
| 1503 | ", NV" | 20 | NV | Remove |
| 1504 | ", N.V" | 1 | NV | Remove |
| 1505 | ", NV." | 3 | NV | Remove |
| 1506 | ", N..V." | 1 | NV | Remove |
| 1507 | " N.V." | 1018 | NV | Remove |
| 1508 | " NV" | 156 | NV | Remove |
| 1509 | " N.V" | 19 | NV | Remove |
| 1510 | " NV." | 4 | NV | Remove |
| 1511 | " N..V" | 0 | NV | Remove |
| 1512 | " N,V." | 1 | NV | Remove |
| 1513 | ", LTD. OY" | 1 | OY | Remove |
| 1514 | " LTD. OY" | 15 | OY | Remove |

| | | | | |
|---|---|---|---|---|
| 1515 | " LTD OY" | 18 | OY | Remove |
| 1517 | " LTD. OY." | 1 | OY | Remove |
| 1518 | " LTD., OY" | 1 | OY | Remove |
| 1519 | " INC. OY" | 3 | OY | Remove |
| 1521 | ", OY." | 1 | OY | Remove |
| 1522 | ", OY" | 13 | OY | Remove |
| 1523 | ", O.Y." | 1 | OY | Remove |
| 1524 | " O.Y." | 1 | OY | Remove |
| 1525 | " OY." | 4 | OY | Remove |
| 1526 | " OY" | 1476 | OY | Remove |
| 1527 | " S.A. SOCIETE ANONYME" | 48 | SA | Remove |
| 1528 | " SA SOCIETE ANONYME" | 4 | SA | Remove |
| 1529 | ", S.A. (SOCIETE ANONYME)" | 2 | SA | Remove |
| 1530 | " S.A. (SOCIETE ANONYME)" | 16 | SA | Remove |
| 1531 | " SA (SOCIETE ANONYME)" | 9 | SA | Remove |
| 1532 | " S.A., SOCIETE ANONYME" | 15 | SA | Remove |
| 1533 | " SA, SOCIETE ANONYME" | 5 | SA | Remove |
| 1534 | " SA, (SOCIETE ANONYME)" | 1 | SA | Remove |
| 1535 | " ET CIE (SOCIETE ANONYME)" | 4 | SA | Replace with " & COMPAGNIE" |
| 1536 | " ET CIE SOCIETE ANONYME" | 8 | SA | Replace with " & COMPAGNIE" |
| 1537 | " ET CIE, SOCIETE ANONYME" | 3 | SA | Replace with " & COMPAGNIE" |
| 1538 | " CIE (SOCIETE ANONYME)" | 7 | SA | Replace with " COMPAGNIE" |
| 1539 | " CIE SOCIETE ANONYME" | 12 | SA | Replace with " COMPAGNIE" |
| 1540 | " CIE. (SOCIETE ANONYME)" | 1 | SA | Replace with " COMPAGNIE" |
| 1541 | " CIE, SOCIETE ANONYME" | 5 | SA | Replace with " COMPAGNIE" |
| 1542 | " CIE. SOCIETE ANONYME" | 3 | SA | Replace with " COMPAGNIE" |
| 1543 | ", SOCIETE ANONYME" | 394 | SA | Remove |
| 1545 | " (SOCIETE ANONYME)" | 309 | SA | Remove |
| 1546 | " SOCIETE ANONYME" | 808 | SA | Remove |
| 1547 | ", SOCIETE, ANONYME" | 1 | SA | Remove |
| 1548 | " (SOCIETE ANONYME" | 3 | SA | Remove |
| 1549 | ", A "SOCIETE ANONYME"" | 1 | SA | Remove |
| 1550 | " ( SOCIETE ANONYME)" | 2 | SA | Remove |
| 1551 | " (FRENCH SOCIETE ANONYME)" | 1 | SA | Remove |
| 1552 | " (A FRENCH SOCIETE ANONYME)" | 1 | SA | Remove |
| 1555 | " A "SOCIETE ANONYME"" | 1 | SA | Remove |
| 1556 | " (SOIETE ANONYME)" | 1 | SA | Remove |
| 1557 | " (STE ANONYME)" | 1 | SA | Remove |
| 1558 | " S.A. SOICIETE ANONYME" | 1 | SA | Remove |
| 1559 | ", SOCIETEY ANONYME" | 1 | SA | Remove |
| 1560 | "(SOCI E/ TE ANONYME)" | 1 | SA | Remove |
| 1561 | "(SOCIET E ANONYME" | 1 | SA | Remove |
| 1562 | "(SOCIETETE ANONYME)" | 2 | SA | Remove |
| 1563 | " (SOCI ET E ANONYME)" | 1 | SA | Remove |
| 1564 | " (SCIETE ANONYME)" | 1 | SA | Remove |
| 1565 | "( SOCIETE ANONYME)" | 1 | SA | Remove |
| 1566 | "(SOCIETE ANONYME)" | 11 | SA | Remove |
| 1567 | "SOCIETE ANONYME" | 12 | SA | Remove |
| 1568 | " MFG. CO. A/S" | 1 | AS | Replace with " MANUFACTURING COMPANY" |
| 1569 | " MFG CO. A/S" | 1 | AS | Replace with " MANUFACTURING COMPANY" |
| 1570 | " CO. A/S" | 22 | AS | Replace with " COMPANY" |
| 1571 | " CO. AS" | 3 | AS | Replace with " COMPANY" |
| 1572 | " CO., A/S" | 2 | AS | Replace with " COMPANY" |
| 1573 | " CO. A./S" | 1 | AS | Replace with " COMPANY" |
| 1574 | ", LTD. A.S." | 1 | AS | Remove |
| 1575 | ", LTD. A/S" | 2 | AS | Remove |
| 1576 | " LTD. A.S" | 1 | AS | Remove |
| 1577 | " LTD. A/S" | 5 | AS | Remove |
| 1578 | " LTD. AS" | 2 | AS | Remove |
| 1579 | " LTD., A/S" | 1 | AS | Remove |
| 1580 | ", A.S." | 45 | AS | Remove |
| 1581 | ", A/S" | 10 | AS | Remove |
| 1582 | ", AS" | 5 | AS | Remove |
| 1583 | ", A.S" | 3 | AS | Remove |
| 1584 | " A/S" | 1950 | AS | Remove |

| | | | | |
|---|---|---|---|---|
| 1585 | " A.S." | 240 | AS | Remove |
| 1586 | " AS" | 794 | AS | Remove |
| 1587 | " A.S" | 38 | AS | Remove |
| 1588 | " A/S/" | 2 | AS | Remove |
| 1589 | " AS." | 6 | AS | Remove |
| 1590 | " A-S" | 5 | AS | Remove |
| 1591 | " A/S." | 2 | AS | Remove |
| 1592 | " A//S" | 1 | AS | Remove |
| 1593 | " /AS" | 1 | AS | Remove |
| 1594 | " CO., INCORPORATED" | 6 | INCORPORATED | Replace with " COMPANY" |
| 1595 | " CO. INCORPORATED" | 2 | INCORPORATED | Replace with " COMPANY" |
| 1596 | ", INCORPORATED." | 1 | INCORPORATED | Remove |
| 1597 | ", INCORPORATED" | 1409 | INCORPORATED | Remove |
| 1598 | " INCORPORATED." | 1 | INCORPORATED | Remove |
| 1599 | " INCORPORATED" | 2251 | INCORPORATED | Remove |
| 1600 | " (INCORPORATED)" | 1 | INCORPORATED | Remove |
| 1601 | " PLC A BRITISH PUBLIC LIMITED COMPANY" | 1 | PLC | Remove |
| 1602 | ", PUBLIC LIMITED COMPANY" | 3 | PLC | Remove |
| 1603 | ", PUBLIC. LIMITED COMPANY" | 1 | PLC | Remove |
| 1604 | " A PUBLIC LIMITED COMPANY" | 1 | PLC | Remove |
| 1605 | " PUBLIC LIMITED COMPANY" | 66 | PLC | Remove |
| 1606 | " LTD., A LIMITED COMPANY" | 1 | PLC | Remove |
| 1607 | ", A LIMITED COMPANY" | 1 | LIMITED | Remove |
| 1608 | " PUBIC LIMITED COMPANY" | 1 | PLC | Remove |
| 1609 | " PUPLIC LIMITED COMPANY" | 1 | PLC | Remove |
| 1610 | " (SARL) LIMITED COMPANY" | 1 | SARL | Remove |
| 1611 | " S.R.L., AN ITALIAN LIMITED COMPANY" | 1 | SRL | Remove |
| 1612 | " N.V. A DUTCH LIMITED COMPANY" | 1 | NV | Remove |
| 1613 | " LIMITED, COMPANY" | 1 | LIMITED | Remove |
| 1614 | " (LIMITED COMPANY)" | 1 | LIMITED | Remove |
| 1615 | " LIMITED COMPANY" | 29 | LIMITED | Remove |
| 1617 | " AND COMPANY" | 78 | | Replace with " & COMPANY" |
| 1618 | " AND COMPANY." | 2 | | Replace with " & COMPANY" |
| 1619 | " MFG., COMPANY" | 1 | | Replace with " MANUFACTURING COMPANY" |
| 1620 | " MFG. COMPANY" | 11 | | Replace with " MANUFACTURING COMPANY" |
| 1621 | ", MFG. CO." | 1 | | Replace with " MANUFACTURING COMPANY" |
| 1622 | " MFG. CO." | 114 | | Replace with " MANUFACTURING COMPANY" |
| 1623 | " MFG. CO" | 3 | | Replace with " MANUFACTURING COMPANY" |
| 1624 | " MFG, CO." | 2 | | Replace with " MANUFACTURING COMPANY" |
| 1625 | " MFG., CO." | 6 | | Replace with " MANUFACTURING COMPANY" |
| 1626 | " M.F.G. CO." | 1 | | Replace with " MANUFACTURING COMPANY" |
| 1627 | " MFG CO." | 5 | | Replace with " MANUFACTURING COMPANY" |
| 1629 | ", LTD. CO." | 7 | LIMITED | Remove |
| 1630 | ", LTD., CO." | 2 | LIMITED | Remove |
| 1631 | " LTD., CO." | 7 | LIMITED | Remove |
| 1632 | " LTD, CO." | 2 | LIMITED | Remove |
| 1633 | " LTD., CO" | 1 | LIMITED | Remove |
| 1634 | " GMBH U. CO." | 28 | GMBH | Replace with " & COMPANY" |
| 1635 | " GMBH U CO." | 2 | GMBH | Replace with " & COMPANY" |
| 1636 | " GMBH U. CO" | 1 | GMBH | Replace with " & COMPANY" |
| 1637 | " GMBH. U. CO" | 1 | GMBH | Replace with " & COMPANY" |
| 1638 | " GMBH. U. CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1639 | " U. CO." | 1 | | Replace with " & COMPANY" |
| 1640 | " GMBH AND CO." | 9 | GMBH | Replace with " & COMPANY" |
| 1641 | " AG AND CO." | 1 | AG | Replace with " & COMPANY" |
| 1642 | " AND CO." | 14 | | Replace with " & COMPANY" |
| 1643 | " CO. (GMBH CO.)" | 1 | GMBH | Replace with " COMPANY" |
| 1644 | " GMBH &CO.." | 1 | GMBH | Replace with " & COMPANY" |
| 1645 | " GMBH +CO." | 5 | GMBH | Replace with " & COMPANY" |
| 1646 | " GMBH+ CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1647 | ", GMBH CO." | 1 | GMBH | Replace with " COMPANY" |
| 1648 | " GMBH CO." | 14 | GMBH | Replace with " COMPANY" |

| | | | | |
|---|---|---|---|---|
| 1649 | " PUBLIC LIMITED CO." | 7 | PLC | Remove |
| 1650 | " N.V. A DUTCH LIMITED CO" | 1 | NV | Remove |
| 1651 | " LIMITED CO." | 5 | LIMITED | Remove |
| 1652 | " LIMITED., CO." | 1 | LIMITED | Remove |
| 1653 | " GMBH UND CO." | 5 | GMBH | Replace with " & COMPANY" |
| 1654 | " GMBH UND CO" | 1 | GMBH | Replace with " & COMPANY" |
| 1655 | " AKTIENGESELLSCHAFT UND CO." | 1 | AG | Replace with " & COMPANY" |
| 1656 | " UND CO." | 2 | | Replace with " & COMPANY" |
| 1657 | " IND., CO." | 1 | | Replace with " INDUSTRAL COMPANY" |
| 1658 | " IND. CO." | 3 | | Replace with "INDUSTRAL COMPANY" |
| 1659 | " AG+ CO." | 1 | AG | Replace with " & COMPANY" |
| 1660 | " AG CO." | 3 | AG | Replace with " COMPANY" |
| 1661 | " INC., CO." | 2 | INCORPORATED | Replace with " COMPANY" |
| 1662 | " INC, CO." | 1 | INCORPORATED | Replace with " COMPANY" |
| 1665 | ", & CO." | 1 | | Replace with " & COMPANY" |
| 1666 | ", AG & CO." | 1 | AG | Replace with " & COMPANY" |
| 1667 | " AG & CO." | 43 | AG | Replace with " & COMPANY" |
| 1668 | " AG & CO" | 8 | AG | Replace with " & COMPANY" |
| 1669 | " AG + CO" | 3 | AG | Replace with " & COMPANY" |
| 1670 | " AG + CO." | 2 | AG | Replace with " & COMPANY" |
| 1671 | " A.G. & CO." | 1 | AG | Replace with " & COMPANY" |
| 1672 | " AKTIENGESELLSCHAFT & CO." | 7 | AG | Replace with " & COMPANY" |
| 1673 | " & CO., GMBH & CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1674 | " & CIE, GMBH & CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1675 | " & CO., (GMBH & CO)" | 1 | GMBH | Replace with " & COMPANY" |
| 1676 | " & CO., (GMBH & CO.)." | 1 | GMBH | Replace with " & COMPANY" |
| 1677 | " & CO., (GMBH & CO.)" | 1 | GMBH | Replace with " & COMPANY" |
| 1678 | " & CO. GMBH & CO." | 4 | GMBH | Replace with " & COMPANY" |
| 1679 | " + CO., GMBH & CO" | 1 | GMBH | Replace with " & COMPANY" |
| 1680 | " & CO. (GMBH & CO)" | 3 | GMBH | Replace with " & COMPANY" |
| 1681 | " & CO. (GMBH & CO.)" | 14 | GMBH | Replace with " & COMPANY" |
| 1682 | " & CO (GMBH & CO.)" | 2 | GMBH | Replace with " & COMPANY" |
| 1683 | " & CO (GMBH & CO)" | 3 | GMBH | Replace with " & COMPANY" |
| 1684 | " & CO. (GMBH) & CO.)" | 1 | GMBH | Replace with " & COMPANY" |
| 1685 | " & CIE. GMBH. & CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1686 | " KG (GMBH & CO.)" | 12 | KG | Replace with " & COMPANY" |
| 1687 | " K.G. (GMBH & CO)" | 1 | KG | Replace with " & COMPANY" |
| 1688 | " KG (GMBH & CO)" | 6 | KG | Replace with " & COMPANY" |
| 1689 | " KG. (GMBH & CO)" | 1 | KG | Replace with " & COMPANY" |
| 1690 | " KG (GMBH) & CO)" | 1 | KG | Replace with " & COMPANY" |
| 1691 | " KG (GMBH + CO.)" | 1 | KG | Replace with " & COMPANY" |
| 1692 | ", GMBH & CO." | 13 | GMBH | Replace with " & COMPANY" |
| 1693 | ", GMBH & CO" | 2 | GMBH | Replace with " & COMPANY" |
| 1694 | " GMBH & CO." | 1937 | GMBH | Replace with " & COMPANY" |
| 1695 | " GMBH & CO" | 193 | GMBH | Replace with " & COMPANY" |
| 1696 | " (GMBH & CO.)" | 26 | GMBH | Replace with " & COMPANY" |
| 1697 | " (GMBH & CO)" | 4 | GMBH | Replace with " & COMPANY" |
| 1698 | " KOMMANDITGES. (GMBH + CO.)" | 1 | KG | Replace with " & COMPANY" |
| 1699 | " GMBH. & CO." | 15 | GMBH | Replace with " & COMPANY" |
| 1700 | " GMBH + CO." | 96 | GMBH | Replace with " & COMPANY" |
| 1701 | " GMBH + CO" | 16 | GMBH | Replace with " & COMPANY" |
| 1702 | " G.M.B.H. & CO." | 4 | GMBH | Replace with " & COMPANY" |
| 1703 | " GMBH. & CO" | 1 | GMBH | Replace with " & COMPANY" |
| 1704 | " GMBH. + CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1705 | " G.M.B.H. & CO" | 1 | GMBH | Replace with " & COMPANY" |
| 1706 | " (GMBH. & CO.)" | 1 | GMBH | Replace with " & COMPANY" |
| 1707 | " GBMH + CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1708 | " GBMH & CO." | 3 | GMBH | Replace with " & COMPANY" |
| 1709 | ", GESELLSCHAFT M.B.H. & CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1710 | " GESELLSCHAFT M.B.H. & CO." | 7 | GMBH | Replace with " & COMPANY" |
| 1711 | " GESELLSCHAFT M.B.H. & CO" | 2 | GMBH | Replace with " & COMPANY" |
| 1712 | " GES. M.B.H. & CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1713 | " GESELLSCHAFT MBH & CO" | 1 | GMBH | Replace with " & COMPANY" |
| 1714 | " GESELLSCHAFT MBH & | 2 | GMBH | Replace with " & COMPANY" |

| | | | | |
|---|---|---|---|---|
| | CO." | | | |
| | " GESELLSCHAFT M.B.H & | | | |
| 1715 | CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1716 | " MBH & CO." | 28 | GMBH | Replace with " & COMPANY" |
| 1717 | " MBH + CO." | 8 | GMBH | Replace with " & COMPANY" |
| 1718 | " MBH. & CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1719 | " M.B.H. & CO." | 1 | GMBH | Replace with " & COMPANY" |
| 1720 | " MBH & CO" | 3 | GMBH | Replace with " & COMPANY" |
| 1721 | " + CO." | 10 | | Replace with " & COMPANY" |
| 1722 | " + CO" | 3 | | Replace with " & COMPANY" |
| 1723 | ", CO." | 127 | | Replace with " COMPANY" |
| 1724 | ", CO" | 4 | | Replace with " COMPANY" |
| 1725 | " CO." | 2843 | | Replace with " COMPANY" |
| 1726 | " CO" | 119 | | Replace with " COMPANY" |
| 1727 | ", MFG. CORP." | 1 | | Replace with " MANUFACTURING CORPORATION" |
| 1728 | " (MFG.) CORP." | 1 | | Replace with " MANUFACTURING CORPORATION" |
| 1729 | " MFG. CORP." | 48 | | Replace with " MANUFACTURING CORPORATION" |
| 1730 | " MFG., CORP." | 7 | | Replace with " MANUFACTURING CORPORATION" |
| 1731 | " MFG CORP." | 2 | | Replace with " MANUFACTURING CORPORATION" |
| 1732 | " MFG. CORP" | 1 | | Replace with " MANUFACTURING CORPORATION" |
| 1733 | " MFG, CORP." | 1 | | Replace with " MANUFACTURING CORPORATION" |
| 1734 | " MFG., CORP" | 1 | | Replace with " MANUFACTURING CORPORATION" |
| 1735 | " MFG CORP" | 1 | | Replace with " MANUFACTURING CORPORATION" |
| 1736 | " INT'L CORP." | 6 | | Replace with " INTERNATIONAL CORPORATION" |
| 1737 | " INT'L. CORP" | 1 | | Replace with " INTERNATIONAL CORPORATION" |
| 1738 | " INT'L. CORP." | 3 | | Replace with " INTERNATIONAL CORPORATION" |
| 1739 | " INTL. CORP." | 1 | | Replace with " INTERNATIONAL CORPORATION" |
| 1740 | ", CORP." | 244 | | Replace with " CORPORATION" |
| 1741 | ", CORP" | 13 | | Replace with " CORPORATION" |
| 1742 | " CORP." | 4638 | | Replace with " CORPORATION" |
| 1743 | " CORP" | 319 | | Replace with " CORPORATION" |
| 1744 | " GMBH CO. KG" | 23 | KG | Replace with " COMPANY" |
| 1745 | " GMBH. CO., KG" | 1 | KG | Replace with " COMPANY" |
| 1746 | " GMBH CO., K.G." | 1 | KG | Replace with " COMPANY" |
| 1747 | " GMBH CO., KG" | 2 | KG | Replace with " COMPANY" |
| 1748 | " GMBH CO, KG" | 1 | KG | Replace with " COMPANY" |
| 1749 | " GMBH +CO. KG" | 4 | KG | Replace with " & COMPANY" |
| 1750 | " GMBH& CO. KG" | 2 | KG | Replace with " & COMPANY" |
| 1751 | " GMBH &CO KG" | 1 | KG | Replace with " & COMPANY" |
| 1752 | " GMBH+ CO. KG" | 2 | KG | Replace with " & COMPANY" |
| 1753 | " GMBH &CO. KG" | 2 | KG | Replace with " & COMPANY" |
| 1754 | " GMBH+ CO KG" | 2 | KG | Replace with " & COMPANY" |
| 1755 | " GMBH +CO KG" | 1 | KG | Replace with " & COMPANY" |
| 1756 | " + CIE., GMBH U. CO. KG" | 1 | KG | Replace with " & COMPANY" |
| | " GESELLSCHAFT M.B.H. U. | | | |
| 1757 | CO. KG" | 2 | KG | Replace with " & COMPANY" |
| 1758 | " GES.M.B.H. U. CO. KG" | 1 | KG | Replace with " & COMPANY" |
| | " GESELLSCHAFT M.B.H U. | | | |
| 1759 | CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1760 | " GMBH U. CO. KG" | 20 | KG | Replace with " & COMPANY" |
| 1761 | " GMBH U. CO. KG." | 4 | KG | Replace with " & COMPANY" |
| 1762 | " GMBH U. CO KG" | 2 | KG | Replace with " & COMPANY" |
| 1763 | " MBH U. CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1764 | " GMBH UND CO. KG" | 10 | KG | Replace with " & COMPANY" |
| 1765 | " GMBH UND CO KG" | 4 | KG | Replace with " & COMPANY" |
| 1766 | " GMBH UND CO. KG." | 2 | KG | Replace with " & COMPANY" |
| 1767 | " M.B.H. UND CO. KG." | 1 | KG | Replace with " & COMPANY" |
| 1768 | " M.B.H. UND CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1769 | " UND CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1770 | " UND CO. KG." | 1 | KG | Replace with " & COMPANY" |
| 1771 | ", GMBH AND CO. KG." | 1 | KG | Replace with " & COMPANY" |
| 1772 | " GMBH AND CO. KG" | 9 | KG | Replace with " & COMPANY" |
| 1773 | " GMBH AND CO., KG" | 1 | KG | Replace with " & COMPANY" |
| 1774 | " AG AND CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1775 | " GMB& O CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1776 | " GES. M.B.H. &CO. KG." | 1 | KG | Replace with " & COMPANY" |
| 1777 | " AG CO. KG" | 1 | KG | Replace with " COMPANY" |
| 1778 | " & CO. GMBH & CO. KG" | 10 | KG | Replace with " & COMPANY" |
| 1779 | " & CO, GMBH & CO KG" | 1 | KG | Replace with " & COMPANY" |
| 1780 | " & GMBH & CO. KG" | 3 | KG | Replace with " & COMPANY" |
| 1781 | ", GMBH & CO. KG" | 14 | KG | Replace with " & COMPANY" |

| | | | | |
|---|---|---|---|---|
| 1782 | ", GMBH & CO, KG" | 1 | KG | Replace with " & COMPANY" |
| 1783 | ", GMBH & CO KG" | 3 | KG | Replace with " & COMPANY" |
| 1784 | ", GMBH & CO. KG." | 1 | KG | Replace with " & COMPANY" |
| 1785 | ", GMBH & CO., KG" | 3 | KG | Replace with " & COMPANY" |
| 1786 | " GMBH & CO. KG" | 4154 | KG | Replace with " & COMPANY" |
| 1787 | " GMBH & CO., KG" | 230 | KG | Replace with " & COMPANY" |
| 1788 | " GMBH & CO KG" | 369 | KG | Replace with " & COMPANY" |
| 1789 | " GMBH & CO. KG." | 268 | KG | Replace with " & COMPANY" |
| 1790 | " GESELLSCHAFT M.B.H. & CO. KG" | 40 | KG | Replace with " & COMPANY" |
| 1791 | " GESELLSCHAFT MBH & CO. KG" | 13 | KG | Replace with " & COMPANY" |
| 1792 | " GES. M.B.H. & CO. KG" | 4 | KG | Replace with " & COMPANY" |
| 1793 | " GMBH. & CO. KG" | 22 | KG | Replace with " & COMPANY" |
| 1794 | " GMBH & CO., KG." | 12 | KG | Replace with " & COMPANY" |
| 1795 | " GMBH & CO. K.G." | 38 | KG | Replace with " & COMPANY" |
| 1796 | " GMBH & CO K.G." | 3 | KG | Replace with " & COMPANY" |
| 1797 | " GMBH & CO KG." | 12 | KG | Replace with " & COMPANY" |
| 1798 | " GESELLSCHAFT MBH & CO., KG" | 1 | KG | Replace with " & COMPANY" |
| 1799 | " GESELLSCHAFT M.B.H. & CO. KG." | 20 | KG | Replace with " & COMPANY" |
| 1800 | " GES.M.B.H. & CO. KG" | 8 | KG | Replace with " & COMPANY" |
| 1801 | " GES.M.B.H & CO. KG" | 3 | KG | Replace with " & COMPANY" |
| 1802 | " GMBH & CO, KG" | 23 | KG | Replace with " & COMPANY" |
| 1804 | " GMBH. & CO., KG" | 2 | KG | Replace with " & COMPANY" |
| 1805 | " G.M.B.H. & CO. KG" | 3 | KG | Replace with " & COMPANY" |
| 1806 | " GES.M.B.H. & CO KG" | 2 | KG | Replace with " & COMPANY" |
| 1807 | " GMBH. & CO. KG." | 13 | KG | Replace with " & COMPANY" |
| 1808 | " & CO. (GMBH & CO. KG)" | 2 | KG | Replace with " & COMPANY" |
| 1809 | " (GMBH & CO.) KG" | 3 | KG | Replace with " & COMPANY" |
| 1810 | " GES.M. B. H & CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1811 | " GNBH & CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1812 | " GES.M.B.H. & CO. KG." | 3 | KG | Replace with " & COMPANY" |
| 1813 | " GMBH & CO, KG." | 1 | KG | Replace with " & COMPANY" |
| 1814 | " GES.MBH & CO. KG" | 3 | KG | Replace with " & COMPANY" |
| 1815 | " GESELLSCHAFT M.B.H. & CO. K.G." | 1 | KG | Replace with " & COMPANY" |
| 1816 | " GES.M.H. & CO KG." | 1 | KG | Replace with " & COMPANY" |
| 1817 | " GESMBH & CO. KG" | 2 | KG | Replace with " & COMPANY" |
| 1818 | " GESELLSCHAFT M.B.H. & CO., KG" | 2 | KG | Replace with " & COMPANY" |
| 1819 | " GESELLSCHAFT M.B.H & CO. KG." | 2 | KG | Replace with " & COMPANY" |
| 1820 | " GMBH & CO., K.G." | 7 | KG | Replace with " & COMPANY" |
| 1821 | " (GMBH & CO. KG)" | 1 | KG | Replace with " & COMPANY" |
| 1822 | " GMGH & CO., KG" | 1 | KG | Replace with " & COMPANY" |
| 1823 | " GESELLSCHAFT MBH & CO KG" | 2 | KG | Replace with " & COMPANY" |
| 1824 | " GMBH & CO: KG" | 2 | KG | Replace with " & COMPANY" |
| 1825 | " GESELLSCHAFT M.B.H & CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1826 | " GBMH & CO. KG" | 2 | KG | Replace with " & COMPANY" |
| 1827 | " (GMBH & CO KG)" | 1 | KG | Replace with " & COMPANY" |
| 1828 | " G.M.B.H & CO., K.G." | 1 | KG | Replace with " & COMPANY" |
| 1829 | " M.B.H. & CO. K.G." | 1 | KG | Replace with " & COMPANY" |
| 1830 | " MBH & CO., KG" | 1 | KG | Replace with " & COMPANY" |
| 1831 | " MBH + CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1832 | " GESELLSCHAFT M.B.H & CO., KG" | 1 | KG | Replace with " & COMPANY" |
| 1833 | " MBH & CO KG" | 3 | KG | Replace with " & COMPANY" |
| 1834 | " M.B.H. & CO KG" | 2 | KG | Replace with " & COMPANY" |
| 1835 | " GMBH + CO KG" | 25 | KG | Replace with " & COMPANY" |
| 1836 | " GMBH + CO. KG." | 9 | KG | Replace with " & COMPANY" |
| 1837 | " GMBH + CO. KG" | 134 | KG | Replace with " & COMPANY" |
| 1838 | " GMBH + CO., KG" | 1 | KG | Replace with " & COMPANY" |
| 1839 | " AG & CO. KG" | 33 | KG | Replace with " & COMPANY" |
| 1840 | " AG & CO., KG" | 3 | KG | Replace with " & COMPANY" |
| 1841 | " AG & CO. K.G." | 1 | KG | Replace with " & COMPANY" |
| 1842 | " A.G. & CO. K.G." | 1 | KG | Replace with " & COMPANY" |
| 1843 | " AG + CO. KG" | 1 | KG | Replace with " & COMPANY" |

| | | | | |
|---|---|---|---|---|
| 1844 | " AKTIENGESELLSCHAFT & CO. KG" | 4 | KG | Replace with " & COMPANY" |
| 1845 | " AG & CO. KG." | 4 | KG | Replace with " & COMPANY" |
| 1846 | " AG & CO KG" | 12 | KG | Replace with " & COMPANY" |
| 1847 | " B.V. & CO. KG" | 7 | KG | Replace with " & COMPANY" |
| 1848 | " KG & CO. KG" | 1 | KG | Replace with " & COMPANY" |
| 1849 | " MBH & CO. KG." | 3 | KG | Replace with " & COMPANY" |
| 1850 | " MBH & CO. KG" | 88 | KG | Replace with " & COMPANY" |
| 1851 | " M.B.H. & CO. KG" | 13 | KG | Replace with " & COMPANY" |
| 1852 | " MBH & CO., KG." | 1 | KG | Replace with " & COMPANY" |
| 1853 | " & CO. KG" | 97 | KG | Replace with " & COMPANY" |
| 1854 | " & CO KG" | 19 | KG | Replace with " & COMPANY" |
| 1855 | " & CO. KG." | 15 | KG | Replace with " & COMPANY" |
| 1856 | " & CO., KG" | 8 | KG | Replace with " & COMPANY" |
| 1857 | " & CO, KG" | 3 | KG | Replace with " & COMPANY" |
| 1858 | " & CO. K.G." | 2 | KG | Replace with " & COMPANY" |
| 1859 | " & CO KG." | 1 | KG | Replace with " & COMPANY" |
| 1860 | " + CO. KG" | 2 | KG | Replace with " & COMPANY" |
| 1861 | " + CO. KG." | 1 | KG | Replace with " & COMPANY" |
| 1862 | " + CO KG" | 2 | KG | Replace with " & COMPANY" |
| 1863 | "-GMBH & CO. KG" | 0 | KG | Replace with "-GESELLSCHAFT & COMPANY" |
| 1864 | " KG" | 1040 | KG | Replace with " KG" |
| 1865 | " KG." | 32 | KG | Replace with " KG" |
| 1866 | " K.G." | 23 | KG | Replace with " KG" |
| 1867 | ", CO., LLC" | 4 | LLC | Replace with " COMPANY" |
| 1868 | ", CO. L.L.C." | 1 | LLC | Replace with " COMPANY" |
| 1869 | " MFG., CO., LLC" | 0 | LLC | Replace with " MANUFACTURING COMPANY" |
| 1870 | " CO., LLC" | 19 | LLC | Replace with " COMPANY" |
| 1871 | " CO., L.L.C." | 11 | LLC | Replace with " COMPANY" |
| 1872 | " CO. LLC" | 16 | LLC | Replace with " COMPANY" |
| 1873 | " CO. L.L.C." | 5 | LLC | Replace with " COMPANY" |
| 1874 | " CO, LLC" | 4 | LLC | Replace with " COMPANY" |
| 1875 | " CO., LLC." | 2 | LLC | Replace with " COMPANY" |
| 1876 | " CO L.L.C." | 1 | LLC | Replace with " COMPANY" |
| 1877 | " CO LLC" | 1 | LLC | Replace with " COMPANY" |
| 1878 | ", INC. LLC" | 3 | LLC | Replace with " COMPANY" |
| 1879 | ", INC., L.L.C." | 1 | LLC | Replace with " COMPANY" |
| 1880 | ", INC., LLC" | 1 | LLC | Replace with " COMPANY" |
| 1881 | ", INC. L.L.C." | 1 | LLC | Replace with " COMPANY" |
| 1882 | " INC, LLC" | 1 | LLC | Replace with " COMPANY" |
| 1883 | " INC., LLC" | 1 | LLC | Replace with " COMPANY" |
| 1884 | " MFG., LLC" | 3 | LLC | Remove |
| 1885 | " MFG., L.L.C." | 1 | LLC | Remove |
| 1886 | " MFG, LLC" | 1 | LLC | Remove |
| 1887 | " MFG, LLC." | 1 | LLC | Remove |
| 1888 | ", LLC" | 3132 | LLC | Remove |
| 1889 | ", LLC." | 272 | LLC | Remove |
| 1890 | ", L.L.C." | 1213 | LLC | Remove |
| 1891 | ", L.L.C" | 20 | LLC | Remove |
| 1892 | ", LL.C." | 3 | LLC | Remove |
| 1893 | ", L.LC." | 1 | LLC | Remove |
| 1894 | ", L.L.C.." | 1 | LLC | Remove |
| 1895 | " L.L.C." | 405 | LLC | Remove |
| 1896 | " LLC." | 68 | LLC | Remove |
| 1897 | " LLC" | 1763 | LLC | Remove |
| 1898 | " (LLC)" | 4 | LLC | Remove |
| 1899 | " L.L.C" | 11 | LLC | Remove |
| 1900 | " L.LC." | 1 | LLC | Remove |
| 1901 | " LL.C" | 1 | LLC | Remove |
| 1902 | " CO. B.V." | 8 | BV | Replace with " COMPANY" |
| 1903 | " CO., B.V." | 2 | BV | Replace with " COMPANY" |
| 1904 | " CO, B.V." | 1 | BV | Replace with " COMPANY" |
| 1905 | " CO B.V." | 1 | BV | Replace with " COMPANY" |
| 1906 | ", B.V." | 174 | BV | Remove |
| 1907 | ", BV" | 14 | BV | Remove |
| 1908 | ", B.V" | 2 | BV | Remove |
| 1909 | " B.V." | 4574 | BV | Remove |
| 1910 | " BV." | 9 | BV | Remove |
| 1911 | " BV" | 340 | BV | Remove |
| 1912 | " B.V" | 36 | BV | Remove |
| 1913 | ", OY AB" | 1 | AB | Remove |

| | | | | |
|---|---|---|---|---|
| 1914 | " OY, AB" | 1 | AB | Remove |
| 1915 | " OY AB" | 65 | AB | Remove |
| 1916 | " CO. AB" | 13 | AB | Replace with " COMPANY" |
| 1917 | " CO AB" | 9 | AB | Replace with " COMPANY" |
| 1918 | " CO A.B." | 1 | AB | Replace with " COMPANY" |
| 1919 | " CO., AB" | 1 | AB | Replace with " COMPANY" |
| 1920 | " AKTIEBOLAG (AB)" | 2 | AB | Remove |
| 1921 | " AKTIEBOLG (AB)" | 1 | AB | Remove |
| 1922 | ", A/B" | 1 | AB | Remove |
| 1923 | ", AB" | 41 | AB | Remove |
| 1924 | ", A.B." | 2 | AB | Remove |
| 1925 | " AB" | 4795 | AB | Remove |
| 1926 | " A.B." | 38 | AB | Remove |
| 1927 | " AB." | 1 | AB | Remove |
| 1928 | " A/B" | 2 | AB | Remove |
| 1929 | " GMBH & CO AG" | 2 | AG | Replace with " & COMPANY" |
| 1930 | " GMBH & CO. AG" | 4 | AG | Replace with " & COMPANY" |
| 1931 | " GMBH & CO., AG" | 1 | AG | Replace with " & COMPANY" |
| 1932 | " AG & CO AG" | 1 | AG | Replace with " & COMPANY" |
| 1933 | " + CO. AG" | 5 | AG | Replace with " & COMPANY" |
| 1934 | " CO. AG" | 67 | AG | Replace with " COMPANY" |
| 1935 | " CO., AG" | 4 | AG | Replace with " COMPANY" |
| 1936 | " CO. AG." | 4 | AG | Replace with " COMPANY" |
| 1937 | " CO AG" | 5 | AG | Replace with " COMPANY" |
| 1938 | " CO. A.G." | 1 | AG | Replace with " COMPANY" |
| 1939 | " CO., A.G." | 1 | AG | Replace with " COMPANY" |
| 1940 | " + CIE AG" | 2 | AG | Replace with " & COMPANY" |
| 1941 | " CIE. AG" | 11 | AG | Replace with " COMPANY" |
| 1942 | " CIE AG" | 13 | AG | Replace with " COMPANY" |
| 1943 | " CIE. A.-G." | 1 | AG | Replace with " COMPANY" |
| 1944 | " CIE. AG." | 2 | AG | Replace with " COMPANY" |
| 1945 | " AKTIENGESELLSCHAFT AG" | 5 | AG | Remove |
| 1946 | " AKTIENGESELLSCHAFT, AG" | 1 | AG | Remove |
| 1947 | ", AG" | 96 | AG | Remove |
| 1948 | ", A.G." | 30 | AG | Remove |
| 1949 | ", A.G" | 1 | AG | Remove |
| 1950 | ", AG." | 5 | AG | Remove |
| 1951 | " AG" | 6156 | AG | Remove |
| 1952 | " A.G." | 246 | AG | Remove |
| 1953 | " AG." | 104 | AG | Remove |
| 1954 | " A.-G." | 3 | AG | Remove |
| 1955 | " (AG)" | 4 | AG | Remove |
| 1956 | " A/G" | 2 | AG | Remove |
| 1957 | " + CO. GESELLSCHAFT MBH" | 1 | GMBH | Replace with " & COMPANY" |
| 1958 | " CO. GESELLSCHAFT M.B.H." | 10 | GMBH | Replace with " COMPANY" |
| 1959 | " CO., GESELLSCHAFT MBH" | 2 | GMBH | Replace with " COMPANY" |
| 1960 | " CO. GESELLSCHAFT MBH" | 2 | GMBH | Replace with " COMPANY" |
| 1961 | " CO., GESELLSCHAFT M.B.H." | 2 | GMBH | Replace with " COMPANY" |
| 1962 | " CO GESELLSCHAFT M.B.H." | 1 | GMBH | Replace with " COMPANY" |
| 1963 | ", GESELLSCHAFT M.B.H." | 2 | GMBH | Remove |
| 1964 | ", GESELLSCHAFT M.B.H" | 1 | GMBH | Remove |
| 1965 | ", GESELLSCHAFT MBH" | 1 | GMBH | Remove |
| 1966 | ", GES. M.B.H." | 1 | GMBH | Remove |
| 1967 | ", GES, M.B.H" | 1 | GMBH | Remove |
| 1968 | " GES. M.B.H." | 40 | GMBH | Remove |
| 1969 | " GES,, M.B.H." | 1 | GMBH | Remove |
| 1970 | " GES M.B.H." | 4 | GMBH | Remove |
| 1971 | " GES. M.B.H" | 3 | GMBH | Remove |
| 1972 | " GES. MBH" | 5 | GMBH | Remove |
| 1973 | " GES M.B.H" | 1 | GMBH | Remove |
| 1974 | " UND CO. MBH" | 1 | GMBH | Replace with " & COMPANY" |
| 1975 | " UND CO MBH" | 1 | GMBH | Replace with " & COMPANY" |
| 1976 | " CO. MBH" | 4 | GMBH | Replace with " COMPANY" |
| 1977 | ", M.B.H." | 5 | GMBH | Remove |
| 1978 | ", MBH" | 6 | GMBH | Remove |
| 1979 | " M.B.H." | 815 | GMBH | Remove |
| 1980 | " MBH" | 1666 | GMBH | Remove |

| | | | | |
|---|---|---|---|---|
| 1981 | " MBH." | 19 | GMBH | Remove |
| 1982 | " M.B.H" | 65 | GMBH | Remove |
| 1983 | " KABUSHIKI KAISHA" | 1914 | KABUSHIKI KAISHA | Remove |
| 1984 | " YUGEN KAISHA" | 24 | KABUSHIKI KAISHA | Remove |
| 1985 | " KABUSIKI KAISHA" | 18 | KABUSHIKI KAISHA | Remove |
| 1986 | " KABUSHIKA KAISHA" | 17 | KABUSHIKI KAISHA | Remove |
| 1987 | " KABUSHI KAISHA" | 13 | KABUSHIKI KAISHA | Remove |
| 1988 | " KUBUSHIKI KAISHA" | 12 | KABUSHIKI KAISHA | Remove |
| 1989 | " KABISHIKI KAISHA" | 11 | KABUSHIKI KAISHA | Remove |
| 1990 | " KABUSHKI KAISHA" | 10 | KABUSHIKI KAISHA | Remove |
| 1991 | " KABUSHIK KAISHA" | 6 | KABUSHIKI KAISHA | Remove |
| 1992 | " KABSHIKI KAISHA" | 6 | KABUSHIKI KAISHA | Remove |
| 1993 | " KAUBSHIKI KAISHA" | 6 | KABUSHIKI KAISHA | Remove |
| 1994 | " KASBUSHIKI KAISHA" | 6 | KABUSHIKI KAISHA | Remove |
| 1995 | " KABUSHIHI KAISHA" | 5 | KABUSHIKI KAISHA | Remove |
| 1996 | " KABUSKIKI KAISHA" | 5 | KABUSHIKI KAISHA | Remove |
| 1997 | " KABSUHIKI KAISHA" | 4 | KABUSHIKI KAISHA | Remove |
| 1998 | " KAUSHIKI KAISHA" | 4 | KABUSHIKI KAISHA | Remove |
| 1999 | " KABUHSIKI KAISHA" | 4 | KABUSHIKI KAISHA | Remove |
| 2000 | " KAUBUSHIKI KAISHA" | 4 | KABUSHIKI KAISHA | Remove |
| 2001 | " KBUSHIKI KAISHA" | 3 | KABUSHIKI KAISHA | Remove |
| 2002 | " KANUSHIKI KAISHA" | 3 | KABUSHIKI KAISHA | Remove |
| 2003 | " KABUSHKIK KAISHA" | 3 | KABUSHIKI KAISHA | Remove |
| 2004 | " KAISHA KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2005 | " KABUSHHIKI KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2006 | " KABUSGIKI KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2007 | " KABUSIHIKI KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2008 | " KABUSBIKI KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2009 | " KABISHUKI KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2010 | " KABUSHIKIA KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2011 | " KABUSHIKO KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2012 | " KABSUSHIKI KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2013 | " YUUGEN KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2014 | " KABUHIKI KAISHA" | 2 | KABUSHIKI KAISHA | Remove |
| 2015 | " KOGYOKABUSHIKI KAISHA" | 2 | KABUSHIKI KAISHA | Replace with " KOGYO" |
| 2016 | " KABUSHILI KAISHA" | 2 | KAISHA | Remove |
| 2017 | " KABUSHUKI KAISHA" | 2 | KABUSHIKI | Remove |

| | | | | |
|---|---|---|---|---|
| | | | KAISHA KABUSHIKI | |
| 2018 | " KAGUSHIKI KAISHA" | 2 | KAISHA KABUSHIKI | Remove |
| 2019 | " KABUSHISKI KAISHA" | 2 | KAISHA KABUSHIKI | Remove |
| 2020 | " KABUBSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2021 | " KABUKSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2022 | " KABURHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2023 | " KABUSAHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2024 | " KABUISHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2025 | " JABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2026 | " AKBUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2027 | " BABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2028 | " BUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2029 | " DABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2030 | " DENKIKABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Replace with " DENKI" |
| 2031 | " FABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2032 | " KABHUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2033 | " HATSUDOKIKABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Replace with " HATSUDOKI" |
| 2034 | " KABSUBSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2035 | " JIDOSHAKABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Replace with " JIDOSHA" |
| 2036 | " JUKOGYOKABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Replace with " JUKOGYO" |
| 2037 | " KABAHIKI KAISHA" | 0 | KAISHA KABUSHIKI | Remove |
| 2038 | " KABASHIKA KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2039 | " KABASIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2040 | " KABBUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2041 | " KABHSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2042 | " KABISHA KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2043 | " KABUSHIKIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2044 | " KAKUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2045 | " KATUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2046 | " KEBUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2047 | " KEBUSKIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2048 | " KAIBUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2049 | " KOGYOOKABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Replace with " KOGYO" |
| 2050 | " KAIBSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2051 | " LABUSHIKI KAISHA" | 1 | KAISHA KABUSHIKI | Remove |
| 2052 | " SHOKUHINKABUSHIKI KAISHA" | 1 | KAISHA | Replace with " SHOKUHIN" |

| | | | | |
|---|---|---|---|---|
| 2053 | " KOGYOLKABUSHIKI KAISHA" | 1 | KABUSHIKI KAISHA | Replace with " KOGYO" |
| 2054 | " KABUSHISHI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2055 | " KABUSHIHIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2056 | " KABUSHIIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2057 | " KABUSHIKE KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2058 | " KABUSHIKHI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2059 | " KABUSHIKII KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2060 | " KABUSHIKIK KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2061 | " KABUSHIKU KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2062 | " KABUSHIMI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2063 | " KAISHUSHIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2064 | " KABUSHINKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2065 | " KABUSHIBI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2066 | " KABUSHKIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2067 | " KABUSHUSHIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2068 | " KABUSIHI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2069 | " KABUSSHIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2070 | " KABUSYIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2071 | " KABUUSHIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2072 | " KABYSHIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2073 | " KAHUSHIKI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |
| 2074 | " KABUSHINI KAISHA" | 1 | KABUSHIKI KAISHA | Remove |

# APPENDIX 3: TOP 200 OCCURRING LAST WORDS

The top 200 occurring last words after data pre-processing, along with the number of names containing the word as a last word, the cumulative number of names for this word and all higher ranked words, and the percentage of the cumulative number of names compared to the total number of names (443,722). Last words are identified on the basis of the last occurrence of a space in a name; then all non-(A-Z) and non-(0-9) characters are removed resulting in a cleaned version of the last word. Used to identify legal form indications occurring at the end of a name.

| | LAST WORD (CLEANED) | NBR OF NAMES | CUM | % | | LAST WORD (CLEANED) | NBR OF NAMES | CUM | % |
|---|---|---|---|---|---|---|---|---|---|
| 1 | INC | 74,949 | 74,949 | 17% | 51 | SYSTEMS | 772 | 288,402 | 65% |
| 2 | LTD | 35,069 | 110,018 | 25% | 52 | TECHNOLOGY | 769 | 289,171 | 65% |
| 3 | LIMITED | 20,459 | 130,477 | 29% | 53 | B | 763 | 289,934 | 65% |
| 4 | GMBH | 17,490 | 147,967 | 33% | 54 | DAVID | 703 | 290,637 | 65% |
| 5 | CORPORATION | 17,348 | 165,315 | 37% | 55 | INDUSTRIES | 674 | 291,311 | 66% |
| 6 | SA | 9,046 | 174,361 | 39% | 56 | TECHNOLOGIES | 673 | 291,984 | 66% |
| 7 | KG | 7,021 | 181,382 | 41% | 57 | AKTIEBOLAG | 670 | 292,654 | 66% |
| 8 | LLC | 6,974 | 188,356 | 42% | 58 | PAUL | 661 | 293,315 | 66% |
| 9 | AG | 6,786 | 195,142 | 44% | 59 | SL | 659 | 293,974 | 66% |
| 10 | SPA | 5,967 | 201,109 | 45% | 60 | JR | 637 | 294,611 | 66% |
| 11 | CO | 5,875 | 206,984 | 47% | 61 | THOMAS | 626 | 295,237 | 67% |
| 12 | COMPANY | 5,806 | 212,790 | 48% | 62 | INSTITUTE | 566 | 295,803 | 67% |
| 13 | SRL | 5,501 | 218,291 | 49% | 63 | PARTNERSHIP | 562 | 296,365 | 67% |
| 14 | | 5,398 | 223,689 | 50% | 64 | T | 560 | 296,925 | 67% |
| 15 | CORP | 5,370 | 229,059 | 52% | 65 | K | 538 | 297,463 | 67% |
| 16 | BV | 5,165 | 234,224 | 53% | 66 | MICHEL | 531 | 297,994 | 67% |
| 17 | AB | 4,979 | 239,203 | 54% | 67 | HANS | 525 | 298,519 | 67% |
| 18 | INCORPORATED | 3,671 | 242,874 | 55% | 68 | JAMES | 521 | 299,040 | 67% |
| 19 | AS | 3,168 | 246,042 | 55% | 69 | SNC | 517 | 299,557 | 68% |
| 20 | MBH | 2,664 | 248,706 | 56% | 70 | JOSEPH | 503 | 300,060 | 68% |
| 21 | A | 2,447 | 251,153 | 57% | 71 | PIERRE | 500 | 300,560 | 68% |
| 22 | DR | 2,321 | 253,474 | 57% | 72 | WILLIAM | 498 | 301,058 | 68% |
| 23 | KAISHA | 2,208 | 255,682 | 58% | 73 | APS | 497 | 301,555 | 68% |
| 24 | J | 2,192 | 257,874 | 58% | 74 | RICHARD | 496 | 302,051 | 68% |
| 25 | ANONYME | 1,698 | 259,572 | 58% | 75 | CHRISTIAN | 457 | 302,508 | 68% |
| 26 | C | 1,693 | 261,265 | 59% | 76 | JEAN | 456 | 302,964 | 68% |
| 27 | M | 1,576 | 262,841 | 59% | 77 | PRODUCTS | 456 | 303,420 | 68% |
| 28 | L | 1,545 | 264,386 | 60% | 78 | MARIA | 450 | 303,870 | 68% |
| 29 | OY | 1,536 | 265,922 | 60% | 79 | WALTER | 446 | 304,316 | 69% |
| 30 | E | 1,446 | 267,368 | 60% | 80 | SAS | 446 | 304,762 | 69% |
| 31 | NV | 1,376 | 268,744 | 61% | 81 | FOUNDATION | 442 | 305,204 | 69% |
| 32 | R | 1,334 | 270,078 | 61% | 82 | MARTIN | 435 | 305,639 | 69% |
| 33 | AKTIENGESELLSCHAFT | 1,174 | 271,252 | 61% | 83 | V | 414 | 306,053 | 69% |
| 34 | W | 1,142 | 272,394 | 61% | 84 | CLAUDE | 412 | 306,465 | 69% |
| 35 | D | 1,120 | 273,514 | 62% | 85 | WOLFGANG | 408 | 306,873 | 69% |
| 36 | PETER | 1,120 | 274,634 | 62% | 86 | JOSEF | 404 | 307,277 | 69% |
| 37 | JOHN | 1,118 | 275,752 | 62% | 87 | EV | 403 | 307,680 | 69% |
| 38 | PLC | 1,077 | 276,829 | 62% | 88 | CHARLES | 400 | 308,080 | 69% |
| 39 | SARL | 1,045 | 277,874 | 63% | 89 | DANIEL | 400 | 308,480 | 70% |
| 40 | H | 1,013 | 278,887 | 63% | 90 | GROUP | 397 | 308,877 | 70% |
| 41 | MICHAEL | 988 | 279,875 | 63% | 91 | RESEARCH | 392 | 309,269 | 70% |
| 42 | DIPLING | 957 | 280,832 | 63% | 92 | FRANCE | 386 | 309,655 | 70% |
| 43 | S | 955 | 281,787 | 64% | 93 | BERNARD | 377 | 310,032 | 70% |
| 44 | G | 936 | 282,723 | 64% | 94 | JACQUES | 372 | 310,404 | 70% |
| 45 | INTERNATIONAL | 865 | 283,588 | 64% | 95 | CENTER | 361 | 310,765 | 70% |
| 46 | LP | 865 | 284,453 | 64% | 96 | WERNER | 360 | 311,125 | 70% |
| 47 | ROBERT | 814 | 285,267 | 64% | 97 | TRUST | 356 | 311,481 | 70% |
| 48 | P | 798 | 286,065 | 64% | 98 | GEORGE | 356 | 311,837 | 70% |
| 49 | UNIVERSITY | 787 | 286,852 | 65% | 99 | DRING | 351 | 312,188 | 70% |
| 50 | F | 778 | 287,630 | 65% | 100 | GERHARD | 343 | 312,531 | 70% |

| | LAST WORD (CLEANED) | NBR OF NAMES | CUM | % | | LAST WORD (CLEANED) | NBR OF NAMES | CUM | % |
|---|---|---|---|---|---|---|---|---|---|
| 101 | KLAUS | 343 | 312,874 | 71% | 151 | MARK | 228 | 326,796 | 74% |
| 102 | HELMUT | 341 | 313,215 | 71% | 152 | I | 227 | 327,023 | 74% |
| 103 | COKG | 339 | 313,554 | 71% | 153 | ROLF | 223 | 327,246 | 74% |
| 104 | ANDRE | 335 | 313,889 | 71% | 154 | KURT | 221 | 327,467 | 74% |
| 105 | ASSOCIATES | 332 | 314,221 | 71% | 155 | HERBERT | 220 | 327,687 | 74% |
| 106 | KK | 331 | 314,552 | 71% | 156 | JEANPIERRE | 218 | 327,905 | 74% |
| 107 | FRANK | 324 | 314,876 | 71% | 157 | SOCIETE | 217 | 328,122 | 74% |
| 108 | N | 323 | 315,199 | 71% | 158 | GESMBH | 211 | 328,333 | 74% |
| 109 | ANTONIO | 320 | 315,519 | 71% | 159 | ARTHUR | 208 | 328,541 | 74% |
| 110 | MANFRED | 316 | 315,835 | 71% | 160 | MARC | 208 | 328,749 | 74% |
| 111 | JAN | 315 | 316,150 | 71% | 161 | ENTERPRISES | 207 | 328,956 | 74% |
| 112 | SERVICES | 314 | 316,464 | 71% | 162 | MARIO | 207 | 329,163 | 74% |
| 113 | PATRICK | 313 | 316,777 | 71% | 163 | LABORATORIES | 206 | 329,369 | 74% |
| 114 | EDWARD | 311 | 317,088 | 71% | 164 | GIOVANNI | 205 | 329,574 | 74% |
| 115 | FRANZ | 305 | 317,393 | 72% | 165 | STEFAN | 204 | 329,778 | 74% |
| 116 | HEINZ | 304 | 317,697 | 72% | 166 | BERND | 204 | 329,982 | 74% |
| 117 | KARL | 303 | 318,000 | 72% | 167 | MED | 204 | 330,186 | 74% |
| 118 | DIETER | 298 | 318,298 | 72% | 168 | VENNOOTSCHAP | 203 | 330,389 | 74% |
| 119 | HAFTUNG | 295 | 318,593 | 72% | 169 | HERMANN | 202 | 330,591 | 75% |
| 120 | PHILIPPE | 293 | 318,886 | 72% | 170 | HEINRICH | 202 | 330,793 | 75% |
| 121 | CIE | 292 | 319,178 | 72% | 171 | ULRICH | 200 | 330,993 | 75% |
| 122 | JURGEN | 285 | 319,463 | 72% | 172 | ROBERTO | 200 | 331,193 | 75% |
| 123 | GEORG | 280 | 319,743 | 72% | 173 | O | 197 | 331,390 | 75% |
| 124 | ANDREAS | 280 | 320,023 | 72% | 174 | MASCHINENFABRIK | 196 | 331,586 | 75% |
| 125 | ALAIN | 278 | 320,301 | 72% | 175 | GUY | 194 | 331,780 | 75% |
| 126 | ANTHONY | 272 | 320,573 | 72% | 176 | DIVISION | 194 | 331,974 | 75% |
| 127 | LIMITEE | 271 | 320,844 | 72% | 177 | JEANCLAUDE | 185 | 332,159 | 75% |
| 128 | OHG | 270 | 321,114 | 72% | 178 | LC | 184 | 332,343 | 75% |
| 129 | JOHANNES | 267 | 321,381 | 72% | 179 | HENRY | 183 | 332,526 | 75% |
| 130 | LTDA | 267 | 321,648 | 72% | 180 | GEORGES | 183 | 332,709 | 75% |
| 131 | AMERICA | 266 | 321,914 | 73% | 181 | FRIEDRICH | 182 | 332,891 | 75% |
| 132 | GERARD | 266 | 322,180 | 73% | 182 | RAYMOND | 179 | 333,070 | 75% |
| 133 | ALEXANDER | 262 | 322,442 | 73% | 183 | USA | 178 | 333,248 | 75% |
| 134 | GIUSEPPE | 261 | 322,703 | 73% | 184 | SYSTEM | 177 | 333,425 | 75% |
| 135 | RUDOLF | 259 | 322,962 | 73% | 185 | NORBERT | 170 | 333,595 | 75% |
| 136 | ALBERT | 258 | 323,220 | 73% | 186 | KARLHEINZ | 170 | 333,765 | 75% |
| 137 | SEISAKUSHO | 251 | 323,471 | 73% | 187 | FRANCIS | 168 | 333,933 | 75% |
| 138 | ROGER | 249 | 323,720 | 73% | 188 | ALBERTO | 167 | 334,100 | 75% |
| 139 | GUNTER | 247 | 323,967 | 73% | 189 | ERNST | 167 | 334,267 | 75% |
| 140 | ENGINEERING | 245 | 324,212 | 73% | 190 | BERNHARD | 165 | 334,432 | 75% |
| 141 | ERIC | 242 | 324,454 | 73% | 191 | HENRI | 165 | 334,597 | 75% |
| 142 | ROLAND | 242 | 324,696 | 73% | 192 | PAOLO | 164 | 334,761 | 75% |
| 143 | BRUNO | 239 | 324,935 | 73% | 193 | LUIGI | 163 | 334,924 | 75% |
| 144 | LOUIS | 238 | 325,173 | 73% | 194 | ANDREW | 162 | 335,086 | 76% |
| 145 | FRANCOIS | 236 | 325,409 | 73% | 195 | JOSE | 162 | 335,248 | 76% |
| 146 | WILHELM | 234 | 325,643 | 73% | 196 | JOACHIM | 161 | 335,409 | 76% |
| 147 | ALFRED | 234 | 325,877 | 73% | 197 | MAURICE | 160 | 335,569 | 76% |
| 148 | RENE | 233 | 326,110 | 73% | 198 | LLP | 160 | 335,729 | 76% |
| 149 | HORST | 230 | 326,340 | 74% | 199 | MARIE | 159 | 335,888 | 76% |
| 150 | CV | 228 | 326,568 | 74% | 200 | ALAN | 158 | 336,046 | 76% |

# APPENDIX 4: TOP 200 OCCURRING FIRST WORDS

The top 200 occurring first words after data pre-processing, together with the number of names containing the word as a first word, the cumulative number of names for this word and all higher ranked words, and the percentage of the cumulative number of names compared to the total number of names (443,722). First words are identified on the basis of the first occurrence of a space in a name, then all non-(A-Z) and non-(0-9) characters are removed resulting in a cleaned version of the first word. Used to identify legal form indications occurring at the beginning of a name.

| | FIRST WORD (CLEANED) | NBR OF NAMES | CUM | % | | FIRST WORD (CLEANED) | NBR OF NAMES | CUM | % |
|---|---|---|---|---|---|---|---|---|---|
| 1 | THE | 5,489 | 5,489 | 1.2% | 51 | LE | 280 | 36,837 | 8.3% |
| 2 | | 5,385 | 10,874 | 2.5% | 52 | NIHON | 275 | 37,112 | 8.4% |
| 3 | SOCIETE | 2,477 | 13,351 | 3.0% | 53 | INNOVATIVE | 274 | 37,386 | 8.4% |
| 4 | KABUSHIKI | 1,293 | 14,644 | 3.3% | 54 | MARTIN | 273 | 37,659 | 8.5% |
| 5 | ADVANCED | 1,076 | 15,720 | 3.5% | 55 | MEDICAL | 271 | 37,930 | 8.5% |
| 6 | AMERICAN | 943 | 16,663 | 3.8% | 56 | LA | 271 | 38,201 | 8.6% |
| 7 | INTERNATIONAL | 936 | 17,599 | 4.0% | 57 | BOARD | 269 | 38,470 | 8.7% |
| 8 | VAN | 923 | 18,522 | 4.2% | 58 | BROWN | 268 | 38,738 | 8.7% |
| 9 | NIPPON | 861 | 19,383 | 4.4% | 59 | THOMAS | 268 | 39,006 | 8.8% |
| 10 | DE | 767 | 20,150 | 4.5% | 60 | INDUSTRIAL | 268 | 39,274 | 8.9% |
| 11 | UNIVERSITY | 686 | 20,836 | 4.7% | 61 | F | 255 | 39,529 | 8.9% |
| 12 | INSTITUT | 615 | 21,451 | 4.8% | 62 | RESEARCH | 253 | 39,782 | 9.0% |
| 13 | HITACHI | 604 | 22,055 | 5.0% | 63 | DEUTSCHE | 250 | 40,032 | 9.0% |
| 14 | NATIONAL | 595 | 22,650 | 5.1% | 64 | VON | 249 | 40,281 | 9.1% |
| 15 | JAPAN | 544 | 23,194 | 5.2% | 65 | OY | 249 | 40,530 | 9.1% |
| 16 | UNITED | 540 | 23,734 | 5.3% | 66 | INSTITUTE | 249 | 40,779 | 9.2% |
| 17 | J | 532 | 24,266 | 5.5% | 67 | MATSUSHITA | 248 | 41,027 | 9.2% |
| 18 | A | 519 | 24,785 | 5.6% | 68 | JOHN | 246 | 41,273 | 9.3% |
| 19 | GENERAL | 515 | 25,300 | 5.7% | 69 | GLOBAL | 241 | 41,514 | 9.4% |
| 20 | NEW | 499 | 25,799 | 5.8% | 70 | PACIFIC | 238 | 41,752 | 9.4% |
| 21 | LEE | 485 | 26,284 | 5.9% | 71 | MULLER | 238 | 41,990 | 9.5% |
| 22 | FIRMA | 472 | 26,756 | 6.0% | 72 | PRECISION | 234 | 42,224 | 9.5% |
| 23 | JOHNSON | 459 | 27,215 | 6.1% | 73 | DIGITAL | 232 | 42,456 | 9.6% |
| 24 | MITSUBISHI | 449 | 27,664 | 6.2% | 74 | ASAHI | 223 | 42,679 | 9.6% |
| 25 | ETABLISSEMENTS | 435 | 28,099 | 6.3% | 75 | SAMSUNG | 222 | 42,901 | 9.7% |
| 26 | SMITH | 433 | 28,532 | 6.4% | 76 | INTEGRATED | 222 | 43,123 | 9.7% |
| 27 | R | 418 | 28,950 | 6.5% | 77 | L | 221 | 43,344 | 9.8% |
| 28 | C | 416 | 29,366 | 6.6% | 78 | MICRO | 216 | 43,560 | 9.8% |
| 29 | COMPAGNIE | 386 | 29,752 | 6.7% | 79 | NORTH | 214 | 43,774 | 9.9% |
| 30 | APPLIED | 386 | 30,138 | 6.8% | 80 | KARL | 210 | 43,984 | 9.9% |
| 31 | SIEMENS | 380 | 30,518 | 6.9% | 81 | POWER | 208 | 44,192 | 10.0% |
| 32 | KIM | 370 | 30,888 | 7.0% | 82 | HER | 206 | 44,398 | 10.0% |
| 33 | DR | 364 | 31,252 | 7.0% | 83 | AIR | 206 | 44,604 | 10.1% |
| 34 | M | 364 | 31,616 | 7.1% | 84 | UNIVERSITE | 205 | 44,809 | 10.1% |
| 35 | AB | 351 | 31,967 | 7.2% | 85 | CREATIVE | 205 | 45,014 | 10.1% |
| 36 | CENTRE | 345 | 32,312 | 7.3% | 86 | SUN | 205 | 45,219 | 10.2% |
| 37 | FUJI | 337 | 32,649 | 7.4% | 87 | WANG | 204 | 45,423 | 10.2% |
| 38 | H | 335 | 32,984 | 7.4% | 88 | ROBERT | 204 | 45,627 | 10.3% |
| 39 | TOKYO | 325 | 33,309 | 7.5% | 89 | LES | 204 | 45,831 | 10.3% |
| 40 | E | 322 | 33,631 | 7.6% | 90 | CHANG | 203 | 46,034 | 10.4% |
| 41 | US | 313 | 33,944 | 7.6% | 91 | TOYO | 203 | 46,237 | 10.4% |
| 42 | ABB | 311 | 34,255 | 7.7% | 92 | T | 199 | 46,436 | 10.5% |
| 43 | B | 298 | 34,553 | 7.8% | 93 | WILLIAMS | 195 | 46,631 | 10.5% |
| 44 | CHEN | 296 | 34,849 | 7.9% | 94 | SCHNEIDER | 195 | 46,826 | 10.6% |
| 45 | G | 289 | 35,138 | 7.9% | 95 | KOREA | 194 | 47,020 | 10.6% |
| 46 | SA | 288 | 35,426 | 8.0% | 96 | OTTO | 193 | 47,213 | 10.6% |
| 47 | S | 287 | 35,713 | 8.0% | 97 | MILLER | 191 | 47,404 | 10.7% |
| 48 | SUMITOMO | 283 | 35,996 | 8.1% | 98 | GEBR | 190 | 47,594 | 10.7% |
| 49 | W | 281 | 36,277 | 8.2% | 99 | NV | 189 | 47,783 | 10.8% |
| 50 | UNIVERSAL | 280 | 36,557 | 8.2% | 100 | SCHMIDT | 189 | 47,972 | 10.8% |

| | FIRST WORD (CLEANED) | NBR OF NAMES | CUM | % | | FIRST WORD (CLEANED) | NBR OF NAMES | CUM | % |
|---|---|---|---|---|---|---|---|---|---|
| 101 | PAUL | 188 | 48,160 | 10.9% | 151 | LG | 140 | 56,185 | 12,7% |
| 102 | LABORATOIRES | 188 | 48,348 | 10.9% | 152 | LASER | 139 | 56,324 | 12,7% |
| 103 | JAMES | 183 | 48,531 | 10.9% | 153 | SILICON | 138 | 56,462 | 12,7% |
| 104 | K | 183 | 48,714 | 11.0% | 154 | GEBRUDER | 136 | 56,598 | 12,8% |
| 105 | ENVIRONMENTAL | 182 | 48,896 | 11.0% | 155 | FIRST | 135 | 56,733 | 12,8% |
| 106 | HANS | 182 | 49,078 | 11.1% | 156 | WALKER | 134 | 56,867 | 12,8% |
| 107 | STATE | 178 | 49,256 | 11.1% | 157 | CONTINENTAL | 134 | 57,001 | 12,8% |
| 108 | TAIWAN | 177 | 49,433 | 11.1% | 158 | PHOENIX | 133 | 57,134 | 12,9% |
| 109 | PARK | 173 | 49,606 | 11.2% | 159 | MEYER | 133 | 57,267 | 12,9% |
| 110 | TAYLOR | 173 | 49,779 | 11.2% | 160 | PARKER | 132 | 57,399 | 12,9% |
| 111 | ALCATEL | 173 | 49,952 | 11.3% | 161 | HEINRICH | 132 | 57,531 | 13,0% |
| 112 | UNION | 172 | 50,124 | 11.3% | 162 | LAIR | 132 | 57,663 | 13,0% |
| 113 | TEXAS | 171 | 50,295 | 11.3% | 163 | HYUNDAI | 130 | 57,793 | 13,0% |
| 114 | WESTERN | 170 | 50,465 | 11.4% | 164 | THOMPSON | 129 | 57,922 | 13,1% |
| 115 | GE | 166 | 50,631 | 11.4% | 165 | BAYER | 129 | 58,051 | 13,1% |
| 116 | P | 165 | 50,796 | 11.4% | 166 | TOSHIBA | 127 | 58,178 | 13,1% |
| 117 | ST | 165 | 50,961 | 11.5% | 167 | AUSTRALIAN | 127 | 58,305 | 13,1% |
| 118 | BRITISH | 165 | 51,126 | 11.5% | 168 | HANSEN | 126 | 58,431 | 13,2% |
| 119 | D | 163 | 51,289 | 11.6% | 169 | BAKER | 126 | 58,557 | 13,2% |
| 120 | WILSON | 162 | 51,451 | 11.6% | 170 | DESIGN | 125 | 58,682 | 13,2% |
| 121 | MASCHINENFABRIK | 162 | 51,613 | 11.6% | 171 | ERNST | 125 | 58,807 | 13,3% |
| 122 | WALTER | 161 | 51,774 | 11.7% | 172 | GREAT | 125 | 58,932 | 13,3% |
| 123 | DELTA | 161 | 51,935 | 11.7% | 173 | CENTRAL | 124 | 59,056 | 13,3% |
| 124 | LIN | 161 | 52,096 | 11.7% | 174 | DAVIS | 124 | 59,180 | 13,3% |
| 125 | KING | 161 | 52,257 | 11.8% | 175 | CAMBRIDGE | 123 | 59,303 | 13,4% |
| 126 | BELL | 161 | 52,418 | 11.8% | 176 | ALPHA | 123 | 59,426 | 13,4% |
| 127 | WORLD | 160 | 52,578 | 11.8% | 177 | BASF | 121 | 59,547 | 13,4% |
| 128 | JONES | 160 | 52,738 | 11.9% | 178 | MOORE | 120 | 59,667 | 13,4% |
| 129 | CARL | 159 | 52,897 | 11.9% | 179 | BAUER | 120 | 59,787 | 13,5% |
| 130 | WILHELM | 158 | 53,055 | 12.0% | 180 | SOUTHERN | 119 | 59,906 | 13,5% |
| 131 | ANDERSON | 158 | 53,213 | 12.0% | 181 | BLUE | 119 | 60,025 | 13,5% |
| 132 | TRW | 154 | 53,367 | 12.0% | 182 | EVANS | 119 | 60,144 | 13,6% |
| 133 | VALEO | 153 | 53,520 | 12.1% | 183 | DOW | 118 | 60,262 | 13,6% |
| 134 | ROYAL | 153 | 53,673 | 12.1% | 184 | TECHNOLOGY | 118 | 60,380 | 13,6% |
| 135 | FRANZ | 153 | 53,826 | 12.1% | 185 | MERCK | 118 | 60,498 | 13,6% |
| 136 | GREEN | 153 | 53,979 | 12.2% | 186 | ETS | 117 | 60,615 | 13,7% |
| 137 | WEBER | 151 | 54,130 | 12.2% | 187 | HARRIS | 117 | 60,732 | 13,7% |
| 138 | WAGNER | 150 | 54,280 | 12.2% | 188 | BECKER | 117 | 60,849 | 13,7% |
| 139 | ENERGY | 150 | 54,430 | 12.3% | 189 | SANYO | 116 | 60,965 | 13,7% |
| 140 | SARL | 149 | 54,579 | 12.3% | 190 | WERNER | 115 | 61,080 | 13,8% |
| 141 | SONY | 149 | 54,728 | 12.3% | 191 | CALIFORNIA | 115 | 61,195 | 13,8% |
| 142 | ELECTRONIC | 149 | 54,877 | 12.4% | 192 | CANADIAN | 115 | 61,310 | 13,8% |
| 143 | DAVID | 149 | 55,026 | 12.4% | 193 | HIGH | 115 | 61,425 | 13,8% |
| 144 | THOMSON | 148 | 55,174 | 12.4% | 194 | LABORATOIRE | 115 | 61,540 | 13,9% |
| 145 | AS | 148 | 55,322 | 12.5% | 195 | CUSTOM | 114 | 61,654 | 13,9% |
| 146 | MITSUI | 147 | 55,469 | 12.5% | 196 | STICHTING | 114 | 61,768 | 13,9% |
| 147 | HONDA | 147 | 55,616 | 12.5% | 197 | 501 | 114 | 61,882 | 13,9% |
| 148 | WHITE | 145 | 55,761 | 12.6% | 198 | CENTRO | 113 | 61,995 | 14,0% |
| 149 | FISCHER | 143 | 55,904 | 12.6% | 199 | DIAMOND | 113 | 62,108 | 14,0% |
| 150 | HUANG | 141 | 56,045 | 12.6% | 200 | PETER | 112 | 62,220 | 14,0% |

# APPENDIX 5: TOP 200 PATENTEES BEFORE NAME CLEANING AND HARMONIZATION

The top 200 patentees based on the original patentee names before name cleaning and harmonization. The number of patents is the sum of all EPO patent applications published between 1978 and 2004 (based on the EPO ESPACE ACCESS product) and all USPTO granted patents published between 1991 and 2003 (based on the USPTO Grant Red Book product).

| RANK | ORIGINAL PATENTEE NAME | PAT | PAT CUM | PAT CUM PCT | PAT EPO | PAT USPTO |
|---|---|---|---|---|---|---|
| 1 | CANON KABUSHIKI KAISHA | 31,649 | 31,649 | 0.98% | 11,293 | 20,356 |
| 2 | SIEMENS AKTIENGESELLSCHAFT | 30,452 | 62,101 | 1.93% | 23,276 | 7,176 |
| 3 | INTERNATIONAL BUSINESS MACHINES CORPORATION | 28,469 | 90,570 | 2.82% | 2,393 | 26,076 |
| 4 | MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD. | 25,594 | 116,164 | 3.61% | 12,576 | 13,018 |
| 5 | SONY CORPORATION | 23,620 | 139,784 | 4.35% | 9,358 | 14,262 |
| 6 | NEC CORPORATION | 23,468 | 163,252 | 5.08% | 7,272 | 16,196 |
| 7 | KABUSHIKI KAISHA TOSHIBA | 23,277 | 186,529 | 5.80% | 8,677 | 14,600 |
| 8 | HITACHI, LTD. | 22,226 | 208,755 | 6.49% | 7,845 | 14,381 |
| 9 | GENERAL ELECTRIC COMPANY | 19,117 | 227,872 | 7.09% | 7,762 | 11,355 |
| 10 | EASTMAN KODAK COMPANY | 18,847 | 246,719 | 7.67% | 7,672 | 11,175 |
| 11 | MITSUBISHI DENKI KABUSHIKI KAISHA | 18,445 | 265,164 | 8.25% | 5,053 | 13,392 |
| 12 | FUJITSU LIMITED | 18,310 | 283,474 | 8.82% | 6,756 | 11,554 |
| 13 | ROBERT BOSCH GMBH | 16,870 | 300,344 | 9.34% | 11,304 | 5,566 |
| 14 | BASF AKTIENGESELLSCHAFT | 16,855 | 317,199 | 9.87% | 11,883 | 4,972 |
| 15 | MOTOROLA, INC. | 15,758 | 332,957 | 10.36% | 4,043 | 11,715 |
| 16 | KONINKLIJKE PHILIPS ELECTRONICS N.V. | 14,411 | 347,368 | 10.80% | 12,374 | 2,037 |
| 17 | SAMSUNG ELECTRONICS CO., LTD. | 13,561 | 360,929 | 11.23% | 3,559 | 10,002 |
| 18 | FUJI PHOTO FILM CO., LTD. | 12,652 | 373,581 | 11.62% | 4,830 | 7,822 |
| 19 | XEROX CORPORATION | 12,104 | 385,685 | 12.00% | 4,104 | 8,000 |
| 20 | INTERNATIONAL BUSINESS MACHINES<BR>CORPORATION | 11,178 | 396,863 | 12.34% | 11,178 | |
| 21 | HEWLETT-PACKARD COMPANY | 11,018 | 407,881 | 12.69% | 3,529 | 7,489 |
| 22 | SHARP KABUSHIKI KAISHA | 10,584 | 418,465 | 13.02% | 4,057 | 6,527 |
| 23 | TEXAS INSTRUMENTS INCORPORATED | 10,353 | 428,818 | 13.34% | 2,962 | 7,391 |
| 24 | BAYER AG | 10,053 | 438,871 | 13.65% | 9,763 | 290 |
| 25 | MINNESOTA MINING AND MANUFACTURING COMPANY | 9,740 | 448,611 | 13.95% | 5,508 | 4,232 |
| 26 | LUCENT TECHNOLOGIES INC. | 9,209 | 457,820 | 14.24% | 3,405 | 5,804 |
| 27 | MICRON TECHNOLOGY, INC. | 9,113 | 466,933 | 14.52% | 338 | 8,775 |
| 28 | HOECHST AKTIENGESELLSCHAFT | 8,596 | 475,529 | 14.79% | 5,914 | 2,682 |
| 29 | INTEL CORPORATION | 8,530 | 484,059 | 15.06% | 1,230 | 7,300 |
| 30 | HONDA GIKEN KOGYO KABUSHIKI KAISHA | 8,224 | 492,283 | 15.31% | 2,928 | 5,296 |
| 31 | SEIKO EPSON CORPORATION | 7,875 | 500,158 | 15.56% | 3,569 | 4,306 |
| 32 | THE PROCTER & GAMBLE COMPANY | 7,300 | 507,458 | 15.78% | 7,300 | |
| 33 | TOYOTA JIDOSHA KABUSHIKI KAISHA | 7,170 | 514,628 | 16.01% | 3,542 | 3,628 |
| 34 | ADVANCED MICRO DEVICES, INC. | 6,799 | 521,427 | 16.22% | 789 | 6,010 |
| 35 | U.S. PHILIPS CORPORATION | 6,389 | 527,816 | 16.42% | | 6,389 |
| 36 | E.I. DU PONT DE NEMOURS AND COMPANY | 6,269 | 534,085 | 16.61% | 6,117 | 152 |
| 37 | PHILIPS ELECTRONICS N.V. | 6,235 | 540,320 | 16.81% | 6,231 | 4 |
| 38 | GENERAL MOTORS CORPORATION | 5,584 | 545,904 | 16.98% | 1,617 | 3,967 |
| 39 | SUN MICROSYSTEMS, INC. | 5,558 | 551,462 | 17.15% | 2,047 | 3,511 |
| 40 | ALCATEL | 5,333 | 556,795 | 17.32% | 4,159 | 1,174 |
| 41 | THE DOW CHEMICAL COMPANY | 5,271 | 562,066 | 17.48% | 2,665 | 2,606 |
| 42 | E. I. DU PONT DE NEMOURS AND COMPANY | 5,229 | 567,295 | 17.65% | 149 | 5,080 |
| 43 | BAYER AKTIENGESELLSCHAFT | 5,219 | 572,514 | 17.81% | 607 | 4,612 |
| 44 | RICOH COMPANY, LTD. | 4,978 | 577,492 | 17.96% | 667 | 4,311 |
| 45 | NOKIA CORPORATION | 4,895 | 582,387 | 18.11% | 4,514 | 381 |
| 46 | L'OREAL | 4,857 | 587,244 | 18.27% | 3,241 | 1,616 |
| 47 | THE REGENTS OF THE UNIVERSITY OF CALIFORNIA | 4,817 | 592,061 | 18.42% | 1,533 | 3,284 |
| 48 | AT&T CORP. | 4,810 | 596,871 | 18.56% | 4,810 | |
| 49 | HENKEL KOMMANDITGESELLSCHAFT AUF AKTIEN | 4,715 | 601,586 | 18.71% | 3,711 | 1,004 |
| 50 | INFINEON TECHNOLOGIES AG | 4,694 | 606,280 | 18.86% | 3,387 | 1,307 |
| 51 | SUMITOMO ELECTRIC INDUSTRIES, LTD. | 4,537 | 610,817 | 19.00% | 2,031 | 2,506 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 52 | NIKON CORPORATION | 4,500 | 615,317 | 19.14% | 753 | 3,747 |
| 53 | KONICA CORPORATION | 4,482 | 619,799 | 19.28% | 2,150 | 2,332 |
| 54 | THE UNITED STATES OF AMERICA AS REPRESENTED BY THE SECRETARY OF THE NAVY | 4,425 | 624,224 | 19.42% | 5 | 4,420 |
| 55 | THE PROCTER & GAMBLE COMPANY | 4,394 | 628,618 | 19.55% | | 4,394 |
| 56 | NISSAN MOTOR CO., LTD. | 4,367 | 632,985 | 19.69% | 1,800 | 2,567 |
| 57 | MICROSOFT CORPORATION | 4,301 | 637,286 | 19.82% | 1,382 | 2,919 |
| 58 | SANYO ELECTRIC CO., LTD. | 4,298 | 641,584 | 19.96% | 1,652 | 2,646 |
| 59 | CIBA-GEIGY AG | 4,068 | 645,652 | 20.08% | 4,049 | 19 |
| 60 | TELEFONAKTIEBOLAGET LM ERICSSON (PUBL) | 4,031 | 649,683 | 20.21% | 2,948 | 1,083 |
| 61 | APPLIED MATERIALS, INC. | 4,023 | 653,706 | 20.33% | 1,508 | 2,515 |
| 62 | ELI LILLY AND COMPANY | 3,990 | 657,696 | 20.46% | 2,213 | 1,777 |
| 63 | 3M INNOVATIVE PROPERTIES COMPANY | 3,935 | 661,631 | 20.58% | 1,713 | 2,222 |
| 64 | HUGHES AIRCRAFT COMPANY | 3,900 | 665,531 | 20.70% | 1,752 | 2,148 |
| 65 | UNILEVER N.V. | 3,856 | 669,387 | 20.82% | 3,854 | 2 |
| 66 | EATON CORPORATION | 3,835 | 673,222 | 20.94% | 1,787 | 2,048 |
| 67 | UNILEVER PLC | 3,716 | 676,938 | 21.06% | 3,713 | 3 |
| 68 | DENSO CORPORATION | 3,674 | 680,612 | 21.17% | 1,301 | 2,373 |
| 69 | MURATA MANUFACTURING CO., LTD. | 3,628 | 684,240 | 21.28% | 1,269 | 2,359 |
| 70 | DAIMLERCHRYSLER AG | 3,599 | 687,839 | 21.39% | 2,074 | 1,525 |
| 71 | YAZAKI CORPORATION | 3,571 | 691,410 | 21.51% | 627 | 2,944 |
| 72 | DELPHI TECHNOLOGIES, INC. | 3,546 | 694,956 | 21.62% | 1,854 | 1,692 |
| 73 | IMPERIAL CHEMICAL INDUSTRIES PLC | 3,514 | 698,470 | 21.73% | 2,387 | 1,127 |
| 74 | OLYMPUS OPTICAL CO., LTD. | 3,448 | 701,918 | 21.83% | 489 | 2,959 |
| 75 | FORD MOTOR COMPANY | 3,401 | 705,319 | 21.94% | 870 | 2,531 |
| 76 | UNITED TECHNOLOGIES CORPORATION | 3,394 | 708,713 | 22.04% | 1,531 | 1,863 |
| 77 | ABBOTT LABORATORIES | 3,365 | 712,078 | 22.15% | 1,527 | 1,838 |
| 78 | THOMSON-CSF | 3,354 | 715,432 | 22.25% | 2,511 | 843 |
| 79 | SUMITOMO CHEMICAL COMPANY, LIMITED | 3,343 | 718,775 | 22.36% | 1,865 | 1,478 |
| 80 | COMMISSARIAT A L'ENERGIE ATOMIQUE | 3,179 | 721,954 | 22.46% | 2,311 | 868 |
| 81 | SHELL INTERNATIONALE RESEARCH<BR>MAATSCHAPPIJ B.V. | 3,164 | 725,118 | 22.55% | 3,164 | |
| 82 | MOBIL OIL CORPORATION | 3,091 | 728,209 | 22.65% | 1,384 | 1,707 |
| 83 | THE BOEING COMPANY | 3,084 | 731,293 | 22.75% | 1,148 | 1,936 |
| 84 | BROTHER KOGYO KABUSHIKI KAISHA | 3,076 | 734,369 | 22.84% | 762 | 2,314 |
| 85 | SUMITOMO WIRING SYSTEMS, LTD. | 3,048 | 737,417 | 22.94% | 1,456 | 1,592 |
| 86 | CATERPILLAR INC. | 2,999 | 740,416 | 23.03% | 789 | 2,210 |
| 87 | NGK INSULATORS, LTD. | 2,995 | 743,411 | 23.12% | 1,649 | 1,346 |
| 88 | KIMBERLY-CLARK WORLDWIDE, INC. | 2,983 | 746,394 | 23.22% | 1,289 | 1,694 |
| 89 | THE WHITAKER CORPORATION | 2,908 | 749,302 | 23.31% | 1,228 | 1,680 |
| 90 | SHIN-ETSU CHEMICAL CO., LTD. | 2,903 | 752,205 | 23.40% | 1,304 | 1,599 |
| 91 | CORNING INCORPORATED | 2,900 | 755,105 | 23.49% | 1,333 | 1,567 |
| 92 | DOW CORNING CORPORATION | 2,892 | 757,997 | 23.58% | 1,619 | 1,273 |
| 93 | BRIDGESTONE CORPORATION | 2,845 | 760,842 | 23.67% | 1,407 | 1,438 |
| 94 | HONEYWELL INC. | 2,824 | 763,666 | 23.75% | 1,419 | 1,405 |
| 95 | BAYERISCHE MOTOREN WERKE AKTIENGESELLSCHAFT | 2,789 | 766,455 | 23.84% | 2,350 | 439 |
| 96 | KAO CORPORATION | 2,725 | 769,180 | 23.92% | 1,545 | 1,180 |
| 97 | SMITHKLINE BEECHAM CORPORATION | 2,723 | 771,903 | 24.01% | 1,506 | 1,217 |
| 98 | PIONEER ELECTRONIC CORPORATION | 2,719 | 774,622 | 24.09% | 989 | 1,730 |
| 99 | LG ELECTRONICS INC. | 2,710 | 777,332 | 24.18% | 1,158 | 1,552 |
| 100 | MEDTRONIC, INC. | 2,658 | 779,990 | 24.26% | 988 | 1,670 |
| 101 | ALLIEDSIGNAL INC. | 2,651 | 782,641 | 24.34% | 1,374 | 1,277 |
| 102 | NORTEL NETWORKS LIMITED | 2,638 | 785,279 | 24.43% | 1,332 | 1,306 |
| 103 | NATIONAL SEMICONDUCTOR CORPORATION | 2,612 | 787,891 | 24.51% | 540 | 2,072 |
| 104 | STMICROELECTRONICS S.R.L. | 2,608 | 790,499 | 24.59% | 1,800 | 808 |
| 105 | ROHM AND HAAS COMPANY | 2,596 | 793,095 | 24.67% | 1,489 | 1,107 |
| 106 | AIR PRODUCTS AND CHEMICALS, INC. | 2,591 | 795,686 | 24.75% | 1,420 | 1,171 |
| 107 | MERCK & CO., INC. | 2,581 | 798,267 | 24.83% | 2,581 | |
| 108 | ASAHI KOGAKU KOGYO KABUSHIKI KAISHA | 2,548 | 800,815 | 24.91% | 102 | 2,446 |
| 109 | LSI LOGIC CORPORATION | 2,536 | 803,351 | 24.99% | 226 | 2,310 |
| 110 | YAMAHA CORPORATION | 2,533 | 805,884 | 25.07% | 600 | 1,933 |
| 111 | FUJI XEROX CO., LTD. | 2,526 | 808,410 | 25.14% | 329 | 2,197 |
| 112 | DIGITAL EQUIPMENT CORPORATION | 2,498 | 810,908 | 25.22% | 722 | 1,776 |
| 113 | PFIZER INC. | 2,490 | 813,398 | 25.30% | 1,337 | 1,153 |
| 114 | TDK CORPORATION | 2,483 | 815,881 | 25.38% | 925 | 1,558 |
| 115 | OKI ELECTRIC INDUSTRY CO., LTD. | 2,480 | 818,361 | 25.45% | 481 | 1,999 |
| 116 | ALPS ELECTRIC CO., LTD. | 2,439 | 820,800 | 25.53% | 894 | 1,545 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 117 | FORD GLOBAL TECHNOLOGIES, INC. | 2,410 | 823,210 | 25.61% | 658 | 1,752 |
| 118 | INSTITUT FRANCAIS DU PETROLE | 2,406 | 825,616 | 25.68% | 1,185 | 1,221 |
| 119 | TRW INC. | 2,389 | 828,005 | 25.75% | 538 | 1,851 |
| 120 | FANUC LTD. | 2,371 | 830,376 | 25.83% | 1,691 | 680 |
| 121 | RAYTHEON COMPANY | 2,355 | 832,731 | 25.90% | 923 | 1,432 |
| 122 | SHELL OIL COMPANY | 2,312 | 835,043 | 25.97% | 129 | 2,183 |
| 123 | WESTINGHOUSE ELECTRIC CORPORATION | 2,243 | 837,286 | 26.04% | 1,696 | 547 |
| 124 | MASSACHUSETTS INSTITUTE OF TECHNOLOGY | 2,240 | 839,526 | 26.11% | 663 | 1,577 |
| 125 | COMPAQ COMPUTER CORPORATION | 2,234 | 841,760 | 26.18% | 446 | 1,788 |
| 126 | TAKEDA CHEMICAL INDUSTRIES, LTD. | 2,233 | 843,993 | 26.25% | 1,313 | 920 |
| 127 | BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY | 2,228 | 846,221 | 26.32% | 1,575 | 653 |
| 128 | THE GOODYEAR TIRE & RUBBER COMPANY | 2,226 | 848,447 | 26.39% | | 2,226 |
| 129 | PHILLIPS PETROLEUM COMPANY | 2,191 | 850,638 | 26.46% | 785 | 1,406 |
| 130 | ERICSSON INC. | 2,170 | 852,808 | 26.53% | 816 | 1,354 |
| 131 | HON HAI PRECISION IND. CO., LTD. | 2,149 | 854,957 | 26.59% | | 2,149 |
| 132 | AT&T CORP. | 2,148 | 857,105 | 26.66% | | 2,148 |
| 133 | AMERICAN CYANAMID COMPANY | 2,147 | 859,252 | 26.73% | 983 | 1,164 |
| 134 | MAZDA MOTOR CORPORATION | 2,133 | 861,385 | 26.79% | 759 | 1,374 |
| 135 | MERCK PATENT GMBH | 2,125 | 863,510 | 26.86% | 2,004 | 121 |
| 136 | WARNER-LAMBERT COMPANY | 2,124 | 865,634 | 26.92% | 1,046 | 1,078 |
| 137 | MERCK & CO., INC. | 2,118 | 867,752 | 26.99% | | 2,118 |
| 138 | NIPPONDENSO CO., LTD. | 2,118 | 869,870 | 27.06% | 395 | 1,723 |
| 139 | CIBA-GEIGY CORPORATION | 2,106 | 871,976 | 27.12% | | 2,106 |
| 140 | BRISTOL-MYERS SQUIBB COMPANY | 2,068 | 874,044 | 27.19% | 1,119 | 949 |
| 141 | MONSANTO COMPANY | 2,068 | 876,112 | 27.25% | 1,068 | 1,000 |
| 142 | THE UNITED STATES OF AMERICA AS REPRESENTED BY THE SECRETARY OF THE ARMY | 2,064 | 878,176 | 27.31% | 6 | 2,058 |
| 143 | MINOLTA CO., LTD. | 2,061 | 880,237 | 27.38% | 100 | 1,961 |
| 144 | ILLINOIS TOOL WORKS INC. | 2,058 | 882,295 | 27.44% | 921 | 1,137 |
| 145 | QUALCOMM INCORPORATED | 2,055 | 884,350 | 27.51% | 1,289 | 766 |
| 146 | INDUSTRIAL TECHNOLOGY RESEARCH INSTITUTE | 2,030 | 886,380 | 27.57% | 64 | 1,966 |
| 147 | NIPPON STEEL CORPORATION | 2,025 | 888,405 | 27.63% | 1,030 | 995 |
| 148 | EXXON RESEARCH AND ENGINEERING COMPANY | 1,994 | 890,399 | 27.69% | 1,274 | 720 |
| 149 | SEIKO INSTRUMENTS INC. | 1,988 | 892,387 | 27.76% | 857 | 1,131 |
| 150 | MOLEX INCORPORATED | 1,979 | 894,366 | 27.82% | 991 | 988 |
| 151 | AT&T BELL LABORATORIES | 1,960 | 896,326 | 27.88% | | 1,960 |
| 152 | NCR CORPORATION | 1,946 | 898,272 | 27.94% | 396 | 1,550 |
| 153 | HYUNDAI ELECTRONICS INDUSTRIES CO., LTD. | 1,941 | 900,213 | 28.00% | 11 | 1,930 |
| 154 | TOKYO ELECTRON LIMITED | 1,924 | 902,137 | 28.06% | 483 | 1,441 |
| 155 | NIPPON TELEGRAPH AND TELEPHONE CORPORATION | 1,898 | 904,035 | 28.12% | 1,051 | 847 |
| 156 | TORAY INDUSTRIES, INC. | 1,857 | 905,892 | 28.18% | 1,193 | 664 |
| 157 | NOVO NORDISK A/S | 1,856 | 907,748 | 28.23% | 939 | 917 |
| 158 | BAYER CORPORATION | 1,835 | 909,583 | 28.29% | 888 | 947 |
| 159 | TELEFONAKTIEBOLAGET LM ERICSSON | 1,834 | 911,417 | 28.35% | 577 | 1,257 |
| 160 | SEMICONDUCTOR ENERGY LABORATORY CO., LTD. | 1,804 | 913,221 | 28.40% | 226 | 1,578 |
| 161 | NOVARTIS AG | 1,791 | 915,012 | 28.46% | 1,309 | 482 |
| 162 | AKZO NOBEL N.V. | 1,785 | 916,797 | 28.52% | 1,275 | 510 |
| 163 | HEWLETT-PACKARD COMPANY,<BR>A DELAWARE CORPORATION | 1,780 | 918,577 | 28.57% | 1,780 | |
| 164 | MITSUBISHI CHEMICAL CORPORATION | 1,780 | 920,357 | 28.63% | 1,036 | 744 |
| 165 | APPLE COMPUTER, INC. | 1,777 | 922,134 | 28.68% | 181 | 1,596 |
| 166 | SCHERING AKTIENGESELLSCHAFT | 1,769 | 923,903 | 28.74% | 1,275 | 494 |
| 167 | UNITED MICROELECTRONICS CORP. | 1,764 | 925,667 | 28.79% | | 1,764 |
| 168 | DEERE & COMPANY | 1,752 | 927,419 | 28.85% | 1,752 | |
| 169 | NORTHROP GRUMMAN CORPORATION | 1,741 | 929,160 | 28.90% | 687 | 1,054 |
| 170 | EASTMAN CHEMICAL COMPANY | 1,737 | 930,897 | 28.95% | 864 | 873 |
| 171 | FRANCE TELECOM | 1,725 | 932,622 | 29.01% | 1,311 | 414 |
| 172 | PITNEY BOWES INC. | 1,717 | 934,339 | 29.06% | 592 | 1,125 |
| 173 | AISIN SEIKI KABUSHIKI KAISHA | 1,707 | 936,046 | 29.11% | 333 | 1,374 |
| 174 | MITSUBISHI HEAVY INDUSTRIES, LTD. | 1,684 | 937,730 | 29.17% | 968 | 716 |
| 175 | YAMAHA HATSUDOKI KABUSHIKI KAISHA | 1,680 | 939,410 | 29.22% | 638 | 1,042 |
| 176 | EXXON CHEMICAL PATENTS INC. | 1,659 | 941,069 | 29.27% | 762 | 897 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 177 | UNISYS CORPORATION | 1,653 | 942,722 | 29.32% | 479 | 1,174 |
| 178 | BAXTER INTERNATIONAL INC. | 1,646 | 944,368 | 29.37% | 776 | 870 |
| 179 | CASIO COMPUTER CO., LTD. | 1,643 | 946,011 | 29.42% | 396 | 1,247 |
| 180 | TEIJIN LIMITED | 1,623 | 947,634 | 29.48% | 1,059 | 564 |
| 181 | DEUTSCHE THOMSON-BRANDT GMBH | 1,620 | 949,254 | 29.53% | 1,265 | 355 |
| 182 | THOMSON LICENSING S.A. | 1,613 | 950,867 | 29.58% | 1,272 | 341 |
| 183 | FORD MOTOR COMPANY LIMITED | 1,611 | 952,478 | 29.63% | 1,611 | |
| 184 | BASF CORPORATION | 1,602 | 954,080 | 29.68% | 634 | 968 |
| 185 | FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER<BR>ANGEWANDTEN FORSCHUNG E.V. | 1,602 | 955,682 | 29.73% | 1,602 | |
| 186 | MITSUI CHEMICALS, INC. | 1,599 | 957,281 | 29.78% | 1,137 | 462 |
| 187 | THE GOODYEAR TIRE & RUBBER COMPANY | 1,592 | 958,873 | 29.82% | 1,592 | |
| 188 | HALLIBURTON ENERGY SERVICES, INC. | 1,579 | 960,452 | 29.87% | 723 | 856 |
| 189 | FORD-WERKE AKTIENGESELLSCHAFT | 1,575 | 962,027 | 29.92% | 1,574 | 1 |
| 190 | BECTON, DICKINSON AND COMPANY | 1,574 | 963,601 | 29.97% | 759 | 815 |
| 191 | EASTMAN KODAK COMPANY<BR>(A NEW JERSEY CORPORATION) | 1,571 | 965,172 | 30.02% | 1,571 | |
| 192 | EBARA CORPORATION | 1,570 | 966,742 | 30.07% | 794 | 776 |
| 193 | HONEYWELL INTERNATIONAL INC. | 1,565 | 968,307 | 30.12% | 718 | 847 |
| 194 | MANNESMANN AKTIENGESELLSCHAFT | 1,556 | 969,863 | 30.17% | 1,055 | 501 |
| 195 | KAWASAKI STEEL CORPORATION | 1,554 | 971,417 | 30.21% | 734 | 820 |
| 196 | SCHERING CORPORATION | 1,554 | 972,971 | 30.26% | 848 | 706 |
| 197 | MITSUBISHI JUKOGYO KABUSHIKI KAISHA | 1,549 | 974,520 | 30.31% | 895 | 654 |
| 198 | HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P. | 1,546 | 976,066 | 30.36% | 496 | 1,050 |
| 199 | STMICROELECTRONICS S.A. | 1,542 | 977,608 | 30.41% | 1,110 | 432 |
| 200 | PIONEER CORPORATION | 1,537 | 979,145 | 30.46% | 1,142 | 395 |

# APPENDIX 6: TOP 200 PATENTEES AFTER NAME CLEANING AND HARMONIZATION

The top 200 patentees based on the patentee names after name cleaning and harmonization. The number of patents is the sum of all EPO patent applications published between 1978 and 2004 (based on the EPO ESPACE ACCESS product) and all USPTO granted patents published between 1991 and 2003 (based on the USPTO Grant Red Book product).

| RANK | HARMONIZED PATENTEE NAME | PAT | PAT CUM | PAT CUM PCT | PAT EPO | PAT USPTO |
|---|---|---|---|---|---|---|
| 1 | INTERNATIONAL BUSINESS MACHINES CORPORATION | 41,173 | 41,173 | 1.28% | 13,575 | 27,598 |
| 2 | CANON | 31,741 | 72,914 | 2.27% | 11,304 | 20,437 |
| 3 | SIEMENS | 30,770 | 103,684 | 3.22% | 23,398 | 7,372 |
| 4 | MATSUSHITA ELECTRIC INDUSTRIAL COMPANY | 26,379 | 130,063 | 4.05% | 12,811 | 13,568 |
| 5 | SONY CORPORATION | 23,665 | 153,728 | 4.78% | 9,358 | 14,307 |
| 6 | NEC CORPORATION | 23,508 | 177,236 | 5.51% | 7,273 | 16,235 |
| 7 | TOSHIBA | 23,344 | 200,580 | 6.24% | 8,696 | 14,648 |
| 8 | HITACHI | 22,754 | 223,334 | 6.95% | 7,934 | 14,820 |
| 9 | GENERAL ELECTRIC COMPANY | 19,620 | 242,954 | 7.56% | 7,762 | 11,858 |
| 10 | EASTMAN KODAK COMPANY | 18,863 | 261,817 | 8.14% | 7,672 | 11,191 |
| 11 | FUJITSU | 18,575 | 280,392 | 8.72% | 6,756 | 11,819 |
| 12 | MITSUBISHI DENKI | 18,513 | 298,905 | 9.30% | 5,053 | 13,460 |
| 13 | BASF | 18,499 | 317,404 | 9.87% | 12,532 | 5,967 |
| 14 | MOTOROLA | 17,294 | 334,698 | 10.41% | 4,401 | 12,893 |
| 15 | BAYER | 17,220 | 351,918 | 10.95% | 11,341 | 5,879 |
| 16 | ROBERT BOSCH | 17,052 | 368,970 | 11.48% | 11,359 | 5,693 |
| 17 | SAMSUNG ELECTRONICS COMPANY | 14,897 | 383,867 | 11.94% | 3,842 | 11,055 |
| 18 | KONINKLIJKE PHILIPS ELECTRONICS | 14,550 | 398,417 | 12.39% | 12,374 | 2,176 |
| 19 | FUJI PHOTO FILM COMPANY | 12,985 | 411,402 | 12.80% | 4,936 | 8,049 |
| 20 | E.I. DU PONT DE NEMOURS & COMPANY | 12,252 | 423,654 | 13.18% | 6,727 | 5,525 |
| 21 | XEROX CORPORATION | 12,111 | 435,765 | 13.55% | 4,104 | 8,007 |
| 22 | HEWLETT-PACKARD COMPANY | 12,024 | 447,789 | 13.93% | 3,747 | 8,277 |
| 23 | THE PROCTER & GAMBLE COMPANY | 11,862 | 459,651 | 14.30% | 7,301 | 4,561 |
| 24 | SHARP | 10,880 | 470,531 | 14.64% | 4,102 | 6,778 |
| 25 | TEXAS INSTRUMENTS | 10,801 | 481,332 | 14.97% | 3,247 | 7,554 |
| 26 | LUCENT TECHNOLOGIES | 10,471 | 491,803 | 15.30% | 3,408 | 7,063 |
| 27 | MINNESOTA MINING AND MANUFACTURING COMPANY | 9,893 | 501,696 | 15.60% | 5,508 | 4,385 |
| 28 | MICRON TECHNOLOGY | 9,320 | 511,016 | 15.89% | 339 | 8,981 |
| 29 | HOECHST | 8,806 | 519,822 | 16.17% | 5,914 | 2,892 |
| 30 | INTEL CORPORATION | 8,560 | 528,382 | 16.43% | 1,230 | 7,330 |
| 31 | HONDA GIKEN KOGYO | 8,264 | 536,646 | 16.69% | 2,930 | 5,334 |
| 32 | ADVANCED MICRO DEVICES | 8,024 | 544,670 | 16.94% | 1,590 | 6,434 |
| 33 | SEIKO EPSON CORPORATION | 7,942 | 552,612 | 17.19% | 3,569 | 4,373 |
| 34 | AT&T CORPORATION | 7,641 | 560,253 | 17.43% | 4,852 | 2,789 |
| 35 | UNILEVER | 7,574 | 567,827 | 17.66% | 7,567 | 7 |
| 36 | TOYOTA JIDOSHA | 7,185 | 575,012 | 17.89% | 3,542 | 3,643 |
| 37 | U.S. PHILIPS CORPORATION | 6,868 | 581,880 | 18.10% | | 6,868 |
| 38 | ALCATEL | 6,679 | 588,559 | 18.31% | 4,898 | 1,781 |
| 39 | CIBA-GEIGY | 6,316 | 594,875 | 18.50% | 4,109 | 2,207 |
| 40 | PHILIPS ELECTRONICS | 6,236 | 601,111 | 18.70% | 6,231 | 5 |
| 41 | STMICROELECTRONICS | 6,224 | 607,335 | 18.89% | 4,065 | 2,159 |
| 42 | RICOH COMPANY | 6,027 | 613,362 | 19.08% | 1,164 | 4,863 |
| 43 | SUMITOMO ELECTRIC INDUSTRIES | 5,968 | 619,330 | 19.26% | 3,317 | 2,651 |
| 44 | SUN MICROSYSTEMS | 5,681 | 625,011 | 19.44% | 2,077 | 3,604 |
| 45 | GENERAL MOTORS CORPORATION | 5,607 | 630,618 | 19.61% | 1,617 | 3,990 |
| 46 | NISSAN MOTOR COMPANY | 5,497 | 636,115 | 19.79% | 2,500 | 2,997 |
| 47 | L'OREAL | 5,490 | 641,605 | 19.96% | 3,314 | 2,176 |
| 48 | THE DOW CHEMICAL COMPANY | 5,292 | 646,897 | 20.12% | 2,665 | 2,627 |
| 49 | FORD MOTOR COMPANY | 5,024 | 651,921 | 20.28% | 2,483 | 2,541 |
| 50 | NOKIA CORPORATION | 4,956 | 656,877 | 20.43% | 4,567 | 389 |
| 51 | MERCK & COMPANY | 4,821 | 661,698 | 20.58% | 2,616 | 2,205 |
| 52 | THE REGENTS OF THE UNIVERSITY OF CALIFORNIA | 4,819 | 666,517 | 20.73% | 1,534 | 3,285 |
| 53 | INFINEON TECHNOLOGIES | 4,793 | 671,310 | 20.88% | 3,391 | 1,402 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 54 | HENKEL KOMMANDITGESELLSCHAFT AUF AKTIEN | 4,715 | 676,025 | 21.03% | 3,711 | 1,004 |
| 55 | TELEFONAKTIEBOLAGET LM ERICSSON (PUBL) | 4,711 | 680,736 | 21.17% | 3,333 | 1,378 |
| 56 | SANYO ELECTRIC COMPANY | 4,654 | 685,390 | 21.32% | 1,828 | 2,826 |
| 57 | SUMITOMO CHEMICAL COMPANY | 4,638 | 690,028 | 21.46% | 2,650 | 1,988 |
| 58 | NIKON CORPORATION | 4,511 | 694,539 | 21.60% | 754 | 3,757 |
| 59 | SMITHKLINE BEECHAM CORPORATION | 4,501 | 699,040 | 21.74% | 2,633 | 1,868 |
| 60 | KONICA CORPORATION | 4,485 | 703,525 | 21.88% | 2,150 | 2,335 |
| 61 | THE UNITED STATES OF AMERICA AS REPRESENTED BY THE SECRETARY OF THE NAVY | 4,443 | 707,968 | 22.02% | 15 | 4,428 |
| 62 | DAIMLERCHRYSLER | 4,398 | 712,366 | 22.16% | 2,151 | 2,247 |
| 63 | MICROSOFT CORPORATION | 4,331 | 716,697 | 22.29% | 1,382 | 2,949 |
| 64 | APPLIED MATERIALS | 4,247 | 720,944 | 22.42% | 1,512 | 2,735 |
| 65 | ELI LILLY & COMPANY | 4,116 | 725,060 | 22.55% | 2,283 | 1,833 |
| 66 | MURATA MANUFACTURING COMPANY | 4,006 | 729,066 | 22.68% | 1,276 | 2,730 |
| 67 | 3M INNOVATIVE PROPERTIES COMPANY | 3,983 | 733,049 | 22.80% | 1,713 | 2,270 |
| 68 | HUGHES AIRCRAFT COMPANY | 3,929 | 736,978 | 22.92% | 1,752 | 2,177 |
| 69 | AGFA-GEVAERT | 3,896 | 740,874 | 23.04% | 2,325 | 1,571 |
| 70 | EATON CORPORATION | 3,857 | 744,731 | 23.16% | 1,809 | 2,048 |
| 71 | THE GOODYEAR TIRE & RUBBER COMPANY | 3,827 | 748,558 | 23.28% | 1,592 | 2,235 |
| 72 | ALLIEDSIGNAL | 3,758 | 752,316 | 23.40% | 1,413 | 2,345 |
| 73 | DENSO CORPORATION | 3,688 | 756,004 | 23.51% | 1,302 | 2,386 |
| 74 | OLYMPUS OPTICAL COMPANY | 3,644 | 759,648 | 23.63% | 501 | 3,143 |
| 75 | DELPHI TECHNOLOGIES | 3,642 | 763,290 | 23.74% | 1,907 | 1,735 |
| 76 | WESTINGHOUSE ELECTRIC CORPORATION | 3,602 | 766,892 | 23.85% | 1,741 | 1,861 |
| 77 | YAZAKI CORPORATION | 3,585 | 770,477 | 23.96% | 627 | 2,958 |
| 78 | IMPERIAL CHEMICAL INDUSTRIES | 3,583 | 774,060 | 24.08% | 2,390 | 1,193 |
| 79 | PFIZER | 3,455 | 777,515 | 24.18% | 1,903 | 1,552 |
| 80 | THOMSON-CSF | 3,437 | 780,952 | 24.29% | 2,536 | 901 |
| 81 | UNITED TECHNOLOGIES CORPORATION | 3,437 | 784,389 | 24.40% | 1,541 | 1,896 |
| 82 | SCHERING | 3,431 | 787,820 | 24.50% | 2,179 | 1,252 |
| 83 | MOBIL OIL CORPORATION | 3,378 | 791,198 | 24.61% | 1,386 | 1,992 |
| 84 | ABBOTT LABORATORIES | 3,367 | 794,565 | 24.71% | 1,529 | 1,838 |
| 85 | NORTEL NETWORKS | 3,353 | 797,918 | 24.82% | 1,547 | 1,806 |
| 86 | CATERPILLAR | 3,332 | 801,250 | 24.92% | 794 | 2,538 |
| 87 | DOW CORNING CORPORATION | 3,254 | 804,504 | 25.02% | 1,833 | 1,421 |
| 88 | COMMISSARIAT A L'ENERGIE ATOMIQUE | 3,237 | 807,741 | 25.12% | 2,352 | 885 |
| 89 | SHELL INTERNATIONALE RESEARCH MAATSCHAPPIJ | 3,234 | 810,975 | 25.22% | 3,185 | 49 |
| 90 | HONEYWELL | 3,201 | 814,176 | 25.32% | 1,635 | 1,566 |
| 91 | LG ELECTRONICS | 3,177 | 817,353 | 25.42% | 1,212 | 1,965 |
| 92 | SUMITOMO WIRING SYSTEMS | 3,149 | 820,502 | 25.52% | 1,472 | 1,677 |
| 93 | FANUC | 3,131 | 823,633 | 25.62% | 2,212 | 919 |
| 94 | SGS-THOMSON MICROELECTRONICS | 3,125 | 826,758 | 25.72% | 844 | 2,281 |
| 95 | THE BOEING COMPANY | 3,098 | 829,856 | 25.81% | 1,148 | 1,950 |
| 96 | BROTHER KOGYO | 3,085 | 832,941 | 25.91% | 762 | 2,323 |
| 97 | SHIN-ETSU CHEMICAL COMPANY | 3,066 | 836,007 | 26.00% | 1,338 | 1,728 |
| 98 | OKI ELECTRIC INDUSTRY COMPANY | 3,053 | 839,060 | 26.10% | 855 | 2,198 |
| 99 | KIMBERLY-CLARK WORLDWIDE | 3,052 | 842,112 | 26.19% | 1,299 | 1,753 |
| 100 | NGK INSULATORS | 3,032 | 845,144 | 26.29% | 1,650 | 1,382 |
| 101 | CORNING | 3,023 | 848,167 | 26.38% | 1,386 | 1,637 |
| 102 | DEERE & COMPANY | 2,990 | 851,157 | 26.47% | 1,752 | 1,238 |
| 103 | DEGUSSA | 2,959 | 854,116 | 26.57% | 2,198 | 761 |
| 104 | BAYERISCHE MOTOREN WERKE | 2,933 | 857,049 | 26.66% | 2,357 | 576 |
| 105 | THE WHITAKER CORPORATION | 2,927 | 859,976 | 26.75% | 1,228 | 1,699 |
| 106 | INSTITUT FRANCAIS DU PETROLE | 2,901 | 862,877 | 26.84% | 1,669 | 1,232 |
| 107 | BRIDGESTONE CORPORATION | 2,873 | 865,750 | 26.93% | 1,421 | 1,452 |
| 108 | FORD GLOBAL TECHNOLOGIES | 2,830 | 868,580 | 27.02% | 848 | 1,982 |
| 109 | MEDTRONIC | 2,807 | 871,387 | 27.10% | 1,026 | 1,781 |
| 110 | PIONEER ELECTRONIC CORPORATION | 2,780 | 874,167 | 27.19% | 989 | 1,791 |
| 111 | NATIONAL SEMICONDUCTOR CORPORATION | 2,751 | 876,918 | 27.28% | 540 | 2,211 |
| 112 | KAO CORPORATION | 2,741 | 879,659 | 27.36% | 1,554 | 1,187 |
| 113 | BECTON, DICKINSON & COMPANY | 2,696 | 882,355 | 27.44% | 1,325 | 1,371 |
| 114 | FUJI XEROX COMPANY | 2,669 | 885,024 | 27.53% | 347 | 2,322 |
| 115 | AIR PRODUCTS AND CHEMICALS | 2,612 | 887,636 | 27.61% | 1,420 | 1,192 |
| 116 | ROHM AND HAAS COMPANY | 2,604 | 890,240 | 27.69% | 1,490 | 1,114 |
| 117 | ERICSSON | 2,590 | 892,830 | 27.77% | 954 | 1,636 |
| 118 | LSI LOGIC CORPORATION | 2,570 | 895,400 | 27.85% | 227 | 2,343 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 119 | TELEFONAKTIEBOLAGET LM ERICSSON | 2,564 | 897,964 | 27.93% | 947 | 1,617 |
| 120 | ASAHI KOGAKU KOGYO | 2,557 | 900,521 | 28.01% | 102 | 2,455 |
| 121 | DIGITAL EQUIPMENT CORPORATION | 2,547 | 903,068 | 28.09% | 722 | 1,825 |
| 122 | YAMAHA CORPORATION | 2,545 | 905,613 | 28.17% | 600 | 1,945 |
| 123 | TAIWAN SEMICONDUCTOR MANUFACTURING COMPANY | 2,508 | 908,121 | 28.25% | 2 | 2,506 |
| 124 | TDK CORPORATION | 2,492 | 910,613 | 28.32% | 925 | 1,567 |
| 125 | ALPS ELECTRIC COMPANY | 2,492 | 913,105 | 28.40% | 894 | 1,598 |
| 126 | UNITED MICROELECTRONICS CORPORATION | 2,450 | 915,555 | 28.48% | 11 | 2,439 |
| 127 | ROHM COMPANY | 2,437 | 917,992 | 28.55% | 963 | 1,474 |
| 128 | COMPAQ COMPUTER CORPORATION | 2,421 | 920,413 | 28.63% | 446 | 1,975 |
| 129 | TRW | 2,421 | 922,834 | 28.70% | 548 | 1,873 |
| 130 | HEIDELBERGER DRUCKMASCHINEN | 2,385 | 925,219 | 28.78% | 1,048 | 1,337 |
| 131 | WARNER-LAMBERT COMPANY | 2,383 | 927,602 | 28.85% | 1,236 | 1,147 |
| 132 | QUALCOMM | 2,379 | 929,981 | 28.93% | 1,378 | 1,001 |
| 133 | RAYTHEON COMPANY | 2,361 | 932,342 | 29.00% | 923 | 1,438 |
| 134 | BRITISH TELECOMMUNICATIONS | 2,358 | 934,700 | 29.07% | 1,598 | 760 |
| 135 | BRISTOL-MYERS SQUIBB COMPANY | 2,342 | 937,042 | 29.15% | 1,172 | 1,170 |
| 136 | SHELL OIL COMPANY | 2,319 | 939,361 | 29.22% | 129 | 2,190 |
| 137 | ILLINOIS TOOL WORKS | 2,316 | 941,677 | 29.29% | 973 | 1,343 |
| 138 | SEAGATE TECHNOLOGY | 2,283 | 943,960 | 29.36% | 181 | 2,102 |
| 139 | TAKEDA CHEMICAL INDUSTRIES | 2,283 | 946,243 | 29.43% | 1,313 | 970 |
| 140 | MASSACHUSETTS INSTITUTE OF TECHNOLOGY | 2,243 | 948,486 | 29.50% | 666 | 1,577 |
| 141 | AMERICAN CYANAMID COMPANY | 2,237 | 950,723 | 29.57% | 983 | 1,254 |
| 142 | HON HAI PRECISION IND. COMPANY | 2,210 | 952,933 | 29.64% | | 2,210 |
| 143 | PHILLIPS PETROLEUM COMPANY | 2,197 | 955,130 | 29.71% | 786 | 1,411 |
| 144 | NIPPONDENSO COMPANY | 2,196 | 957,326 | 29.78% | 396 | 1,800 |
| 145 | HYUNDAI ELECTRONICS INDUSTRIES COMPANY | 2,194 | 959,520 | 29.84% | 12 | 2,182 |
| 146 | MERCK PATENT | 2,188 | 961,708 | 29.91% | 2,004 | 184 |
| 147 | MONSANTO COMPANY | 2,175 | 963,883 | 29.98% | 1,078 | 1,097 |
| 148 | BLACK & DECKER | 2,174 | 966,057 | 30.05% | 857 | 1,317 |
| 149 | EXXON RESEARCH AND ENGINEERING COMPANY | 2,156 | 968,213 | 30.12% | 1,274 | 882 |
| 150 | MAZDA MOTOR CORPORATION | 2,156 | 970,369 | 30.18% | 759 | 1,397 |
| 151 | FRAUNHOFER-GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG E.V. | 2,155 | 972,524 | 30.25% | 1,760 | 395 |
| 152 | AKZO NOBEL | 2,125 | 974,649 | 30.32% | 1,291 | 834 |
| 153 | COLGATE-PALMOLIVE COMPANY | 2,124 | 976,773 | 30.38% | 651 | 1,473 |
| 154 | MINOLTA COMPANY | 2,112 | 978,885 | 30.45% | 100 | 2,012 |
| 155 | F. HOFFMANN-LA ROCHE | 2,090 | 980,975 | 30.51% | 2,080 | 10 |
| 156 | THE UNITED STATES OF AMERICA AS REPRESENTED BY THE SECRETARY OF THE ARMY | 2,080 | 983,055 | 30.58% | 8 | 2,072 |
| 157 | SEIKO INSTRUMENTS | 2,078 | 985,133 | 30.64% | 861 | 1,217 |
| 158 | TOKYO ELECTRON | 2,067 | 987,200 | 30.71% | 489 | 1,578 |
| 159 | NIPPON TELEGRAPH AND TELEPHONE CORPORATION | 2,064 | 989,264 | 30.77% | 1,132 | 932 |
| 160 | VOLKSWAGEN | 2,051 | 991,315 | 30.83% | 1,667 | 384 |
| 161 | INDUSTRIAL TECHNOLOGY RESEARCH INSTITUTE | 2,039 | 993,354 | 30.90% | 64 | 1,975 |
| 162 | NIPPON STEEL CORPORATION | 2,039 | 995,393 | 30.96% | 1,030 | 1,009 |
| 163 | SUMITOMO RUBBER INDUSTRIES | 2,018 | 997,411 | 31.02% | 955 | 1,063 |
| 164 | MOLEX | 1,986 | 999,397 | 31.09% | 993 | 993 |
| 165 | AT&T BELL LABORATORIES | 1,964 | 1,001,361 | 31.15% | 1 | 1,963 |
| 166 | NOVARTIS | 1,957 | 1,003,318 | 31.21% | 1,309 | 648 |
| 167 | EXXON CHEMICAL PATENTS | 1,952 | 1,005,270 | 31.27% | 762 | 1,190 |
| 168 | NCR CORPORATION | 1,949 | 1,007,219 | 31.33% | 396 | 1,553 |
| 169 | PITNEY BOWES | 1,918 | 1,009,137 | 31.39% | 747 | 1,171 |
| 170 | SEMICONDUCTOR ENERGY LABORATORY COMPANY | 1,905 | 1,011,042 | 31.45% | 226 | 1,679 |
| 171 | ASEA BROWN BOVERI | 1,899 | 1,012,941 | 31.51% | 1,061 | 838 |
| 172 | LUCAS INDUSTRIES | 1,894 | 1,014,835 | 31.57% | 1,269 | 625 |
| 173 | TORAY INDUSTRIES | 1,888 | 1,016,723 | 31.62% | 1,194 | 694 |
| 174 | HONEYWELL INTERNATIONAL | 1,887 | 1,018,610 | 31.68% | 871 | 1,016 |
| 175 | HEWLETT-PACKARD COMPANY, A DELAWARE CORPORATION | 1,887 | 1,020,497 | 31.74% | 1,887 | |
| 176 | NOVO NORDISK | 1,872 | 1,022,369 | 31.80% | 939 | 933 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 177 | CASIO COMPUTER COMPANY | 1,848 | 1,024,217 | 31.86% | 589 | 1,259 |
| 178 | FRANCE TELECOM | 1,824 | 1,026,041 | 31.91% | 1,387 | 437 |
| 179 | DR.ING.H.C. F. PORSCHE | 1,811 | 1,027,852 | 31.97% | 1,151 | 660 |
| 180 | APPLE COMPUTER | 1,808 | 1,029,660 | 32.03% | 187 | 1,621 |
| 181 | DAIKIN INDUSTRIES | 1,804 | 1,031,464 | 32.08% | 1,124 | 680 |
| 182 | MITSUBISHI CHEMICAL CORPORATION | 1,798 | 1,033,262 | 32.14% | 1,037 | 761 |
| 183 | MITA INDUSTRIAL COMPANY | 1,788 | 1,035,050 | 32.19% | 841 | 947 |
| 184 | VICTOR COMPANY OF JAPAN | 1,785 | 1,036,835 | 32.25% | 858 | 927 |
| 185 | UNISYS CORPORATION | 1,778 | 1,038,613 | 32.30% | 483 | 1,295 |
| 186 | ASAHI GLASS COMPANY | 1,773 | 1,040,386 | 32.36% | 966 | 807 |
| 187 | NORTHROP GRUMMAN CORPORATION | 1,769 | 1,042,155 | 32.42% | 692 | 1,077 |
| 188 | BAXTER INTERNATIONAL | 1,766 | 1,043,921 | 32.47% | 776 | 990 |
| 189 | HUGHES ELECTRONICS CORPORATION | 1,762 | 1,045,683 | 32.52% | 526 | 1,236 |
| 190 | MANNESMANN | 1,760 | 1,047,443 | 32.58% | 1,057 | 703 |
| 191 | EASTMAN CHEMICAL COMPANY | 1,748 | 1,049,191 | 32.63% | 864 | 884 |
| 192 | NOKIA MOBILE PHONES | 1,748 | 1,050,939 | 32.69% | 217 | 1,531 |
| 193 | MITSUBISHI HEAVY INDUSTRIES | 1,726 | 1,052,665 | 32.74% | 974 | 752 |
| 194 | TEKTRONIX | 1,726 | 1,054,391 | 32.80% | 881 | 845 |
| 195 | UOP | 1,726 | 1,056,117 | 32.85% | 497 | 1,229 |
| 196 | AISIN SEIKI | 1,722 | 1,057,839 | 32.90% | 335 | 1,387 |
| 197 | THOMSON LICENSING | 1,718 | 1,059,557 | 32.96% | 1,282 | 436 |
| 198 | YAMAHA HATSUDOKI | 1,694 | 1,061,251 | 33.01% | 638 | 1,056 |
| 199 | SCHLUMBERGER TECHNOLOGY CORPORATION | 1,694 | 1,062,945 | 33.06% | 434 | 1,260 |
| 200 | HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P. | 1,689 | 1,064,634 | 33.11% | 496 | 1,193 |

# APPENDIX 7: VALIDATION EXERCICE ON 35 HARMONIZED NAMES

A detailed validation exercise was conducted for 35 harmonized names to check for accuracy and completeness: 10 harmonized names that were matched to 1 original patentee name, 5 harmonized names matched to 2 original patentee names, 10 names matched to 3 to 9 original patentee names, and 10 harmonized names matched with 10 or more original patentee names.

Table 35 contains the results of the validation. The second column contains the number of original patentee names automatically identified by the name cleaning and harmonization method. The third column contains the number of patents assigned to these original patentee names. The fourth and the fifth column contain the number of original patentee names and assigned number of patents after validation respectively. The one but last column contains the number of patents assigned to the manually found additional name variations. The last column contains the difference in number of patents compared to the total number patents assigned to all manually identified name variations (cases in bold indicate completeness below 97%)[16].

**Table 35: Comparison of automatically and manually identified name variations for 35 harmonized names**

| HARMONIZED NAME | NBR NAMES | NBR PAT | NBR NAMES (VAL) | NBR PAT (VAL) | DIFF | PCT |
|---|---|---|---|---|---|---|
| MATSUSHITA ELECTRIC INDUSTRIAL COMPANY | 38 | 26,379 | 107 | 26,576 | 69 | 0.26% |
| SOCIETE NATIONALE D'ETUDE ET DE CONSTRUCTION DE MOTEURS D'AVIATION, "S.N.E.C.M.A." | 29 | 956 | 86 | 1,092 | 57 | **5.22%** |
| AT&T CORPORATION | 26 | 7,641 | 31 | 7,668 | 5 | 0.07% |
| SANYO ELECTRIC COMPANY | 24 | 4,654 | 35 | 4,665 | 11 | 0.24% |
| HONDA GIKEN KOGYO | 23 | 8,264 | 84 | 8,473 | 61 | 0.72% |
| SANSHIN KOGYO | 14 | 811 | 21 | 820 | 7 | 0.85% |
| PETROLEO BRASILEIRO S.A. - PETROBRAS | 13 | 174 | 27 | 215 | 14 | **6.51%** |
| ETHICON ENDO-SURGERY | 11 | 622 | 16 | 628 | 6 | 0.96% |
| CENTRAL GLASS COMPANY | 10 | 457 | 10 | 457 | 0 | 0.00% |
| NIPPON ZEON COMPANY | 10 | 669 | 12 | 671 | 2 | 0.30% |
| MEDTRONIC | 9 | 2,807 | 9 | 2,807 | 0 | 0.00% |
| BASF | 8 | 18,499 | 43 | 18,577 | 35 | 0.19% |
| VAN DEN BERGH FOODS CO., DIVISION OF CONOPCO | 8 | 128 | 12 | 154 | 4 | 2.60% |
| TECHNOLOGICAL RESOURCES | 7 | 105 | 9 | 107 | 2 | 1.87% |
| SMITHKLINE BEECHAM BIOLOGICALS | 6 | 101 | 8 | 103 | 2 | 1.94% |
| WAKO PURE CHEMICAL INDUSTRIES | 5 | 309 | 7 | 315 | 2 | 0.63% |
| THOMAS INDUSTRIES | 4 | 114 | 4 | 114 | 0 | 0.00% |
| M.A. INDUSTRIES | 3 | 10 | 3 | 10 | 0 | 0.00% |
| COMMUNICATIONS TECHNOLOGY CORPORATION | 3 | 23 | 3 | 23 | 0 | 0.00% |
| DATABOOK | 3 | 5 | 3 | 5 | 0 | 0.00% |
| VENTREX LABORATORIES | 2 | 6 | 2 | 6 | 0 | 0.00% |
| JOSLIN DIABETES CENTER | 2 | 22 | 4 | 28 | 2 | **7.14%** |
| LICENCIA HOLDING | 2 | 3 | 2 | 3 | 0 | 0.00% |
| MIPS COMPANY | 2 | 7 | 2 | 7 | 0 | 0.00% |
| SANYO KIKI | 2 | 3 | 2 | 3 | 0 | 0.00% |
| PI-PATENT | 1 | 1 | 2 | 2 | 1 | **50.00%** |
| PERFECT PUTT | 1 | 2 | 1 | 2 | 0 | 0.00% |
| RAINTREE ESSIX | 1 | 2 | 1 | 2 | 0 | 0.00% |
| SHEETS ELECTRONICS | 1 | 1 | 1 | 1 | 0 | 0.00% |
| SILENTOR | 1 | 4 | 1 | 4 | 0 | 0.00% |
| COMPOSITE ROTOR | 1 | 5 | 2 | 8 | 1 | **12.50%** |
| NISSHO GIKEN | 1 | 3 | 1 | 3 | 0 | 0.00% |

[16] The same dataset was used as the one for the application of the method (all 270,635 applicant names from all 1,600,812 EPO patent applications published between 1978 and 2004 and all 223,665 assignee names from all 1,614,224 USPTO granted patents published between 1991 and 2003.

| | | | | | | |
|---|---|---|---|---|---|---|
| DIMPLEX NORTH AMERICA | 1 | 22 | 1 | 22 | 0 | 0.00% |
| ADIR ET COMPAGNIE | 1 | 615 | 4 | 639 | 3 | 0.47% |
| YOKOGAWA-HEWLETT-PACKARD | 1 | 1 | 1 | 1 | 0 | 0.00% |
| | **274** | **73,425** | **557** | **74,211** | **284** | **0.38%** |

Overall, 74,211 patens were assigned to all 557 manually identified spelling variations, but only 284 (0.38%) of them are assigned to name variations not found by the automatic procedure. This results in a completeness of 99,62% and an accuracy of 100% as no mismatches were found.

The completeness on the level of the patentee varies between 50% and 100%; 5 cases have coverage below 97%, but only one has coverage below 85%. This one example ("PI-PATENT") with 50% coverage is due to the identification of one additional name variant with one patent.