

Visual Exploration of Patent Collections with IPC Clouds

Dominik Herr^{1,2}, Qi Han¹, Steffen Lohmann¹, Sören Brüggemann³, Thomas Ertl¹

¹Institute for Visualization and Interactive Systems (VIS)

²Graduate School of Excellence advanced Manufacturing Engineering (GSaME)

University of Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany

{dominik.herr, qi.han, steffen.lohmann, thomas.ertl}@vis.uni-stuttgart.de

³Brüggemann Software

Bokeler Straße 18, 26871 Papenburg, Germany

sb@brueggemann-software.eu

ABSTRACT

The International Patent Classification (IPC) is the most widely used system for the classification of patents. It is indispensable in patent retrieval, as it allows to filter patents by their IPC classes, groups, and subgroups. However, the selection of appropriate IPC symbols can be challenging and there is the risk that important patents are overlooked because relevant IPC symbols are not considered in the search. Therefore, the identification of appropriate IPC symbols is a crucial activity in patent retrieval that could significantly benefit from better IT support. This paper introduces IPC clouds, an interactive visualization technique that shows the relatedness of IPC symbols based on their co-use in the patent data. In contrast to the IPC hierarchy, IPC clouds allow to dynamically explore the IPC space while taking into account how the IPC symbols are actually used in the patent data. They provide an alternative view on the IPC system and assist in identifying relevant IPC symbols and associated patents. The general visualization technique is not limited to the IPC system but can also be applied to similar classification systems or to keywords and concepts extracted from the patent documents.

Categories and Subject Descriptors

H.2.8 [Information interfaces and presentation]: User Interfaces—*Graphical user interfaces (GUI)*

Keywords

Patents, retrieval, mining, IPC, CPC, classification, visual analysis, tag cloud, visualization.

1. INTRODUCTION

A technological advantage over competitors is often the key to a superior positioning on the market in today's industry. Therefore, the protection of intellectual property becomes

more and more important. At the same time, it is important to know what the relevant patents in a certain field are. As more than one million patents are issued each year [13], it is increasingly challenging to find the relevant ones.

The International Patent Classification (IPC) is “one of the most important tools available to people who want to search patent databases” [7]. It is developed and maintained by the World Intellectual Property Organization (WIPO) for more than 40 years and used by almost all patent offices for the classification of patents. The IPC divides technology into eight thematic sections with more than 70,000 subdivisions that are hierarchically organized. The IPC symbols are usually assigned to the patents by the national offices that publish the patent documents.

The IPC system can be very useful in navigating the patent database and retrieving relevant patents. Its hierarchical structure allows to filter patents by their IPC classes, subclasses, groups, or subgroups. Often, a set of IPC symbols is used to retrieve patterns of interest for a deeper analysis. This bears the risk that relevant patents are not considered only because they are classified with other IPC symbols than expected. An overview on the actual use and particularly the co-use of IPC symbols would therefore be most helpful to discover related IPC symbols that could be relevant in a certain retrieval context. Inspired by the tag cloud visualization technique [23], we developed *IPC clouds* to visualize the co-use of IPC symbols in patent data and to support the identification of relevant relationships within the IPC space. IPC symbols that are identified to be related can be from very different classes or groups of the IPC hierarchy but may fruitfully extend the set of IPC symbols already used in patent retrieval.

In this paper, we introduce IPC clouds in detail and describe their creation from patent data. Our implementation uses a noSQL database containing bibliographic data for a large amount of patents. We first compute the similarities of each pair of IPC symbols based on their co-use in the patent documents. We then map the similarities on a two-dimensional plane to get a global representation of the IPC space. Based on this mapping, we developed two different types of IPC clouds, one giving a general overview on the IPC space and another focusing on selected IPC symbols. Both visualizations offer several interaction techniques to further support the exploration of the IPC space.

Copyright © 2014 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors. Published at CEUR-WS.org

Proceedings of the 1st International Workshop on Patent Mining and Its Applications (IPaMin 2014). Hildesheim, Oct. 7, 2014. At KONVENS 2014, Oct. 8–10, 2014, Hildesheim, Germany.

2. RELATED WORK

Modern systems for patent retrieval and analysis increasingly provide interactive visualizations to improve access to patent data. As an example, PatAnalyse [10] shows weighted links between applicants and other patent data in matrix visualizations with histograms and color scales. The patent documents themselves are often represented as high dimensional data objects using vector space models. Examples are the “landscape maps” in Patent iNSIGHT Pro [11] or the ThemeScape maps in Thomson Aureka [12].

Another popular visualization technique in the patent domain are node-link diagrams. They are often used in patent citation analysis [16, 21] to show relationships between patents based on citation links. A commercial system incorporating such node-link diagrams is Delphion Citation Link [1]. Other approaches use node-link diagrams to show relations between patents and priority documents [15], or to graphically depict networks of applicants or inventors [21]. Node-link diagrams can be very useful to explore the patent space and to identify important clusters in the patent data.

The IPC space is rarely visualized in related work. Usually, it is shown in some kind of tree view that the user can navigate to find IPC symbols of interest. Kutz uses a sequence of treemaps to visualize the evolution of the IPC system over time [17]. However, the treemaps are again structured according to the IPC hierarchy without considering other IPC relations in the patent data.

IPC clouds, in contrast, do not make use of the IPC hierarchy but visualize the relatedness of IPC symbols based on their actual co-use in the patent data. Furthermore, the IPC relatedness is not explicitly visualized but implicitly by their spatial arrangement, similar to the idea of clustered tag clouds [18]. Also, like in tag clouds, the labels are weighted in the visualization so that their size reflects the usage frequency of the corresponding IPC symbol.

3. PATENT DATA

We use the document-oriented NoSQL database Elastic Search [2] to store the patent data. A document-oriented database has some advantages over a relational one in text mining contexts. In particular, it is less rigid than a relational database in that it does not require a certain data schema or a clear structuring for every record. Different records can have different fields and semi-structured data is usually not a problem. New information can easily be added to a subset of records without the need to update other records in the database or to use empty fields.

Another useful characteristic of document-oriented databases is that they typically allow to retrieve documents based on their content. Elastic Search is based on Apache Lucene, which is a powerful text search engine offering sophisticated full-text indexing and searching. Both Elastic Search and Apache Lucene are open source projects written in Java and released under the Apache License. The patent data is accessible via HTTP and exchanged in JSON format, i.e., it can be retrieved over the web via a RESTful web service. Moreover, we can directly access the Lucene repository to preprocess the data and perform computationally expensive tasks, such as the later described computation of similarities.

The database comprises two repositories, a large one with bibliographic information and a smaller one containing the texts from the patent documents. The bibliographic information was taken from the PatStat database [5] of the European Patent Office. It includes the patent ID, title, abstract, applicant, inventor, filing and application dates, IPC symbols, as well as citations for more than 70 million patents. We transformed the PatStat data into the JSON structure of our Elastic Search database using MongoDB [8].

The patent texts comprise the descriptions and claims for 88,000 arbitrarily chosen patents. They were retrieved from Espacenet [3], the European Patent Register [6], and the European Publication Server [4], using RESTful web services of the Open Patent Services [9]. All texts are indexed by Lucene and linked to the bibliographic information via their unique patent IDs. In this paper, we will focus on how the IPC symbols are used in the patent data.

4. DATA PREPROCESSING

Before IPC clouds are generated, the patent data is pre-processed. The preprocessing consists of two steps: We first compute the pairwise similarities between the IPC symbols and then map these similarities onto a 2D space.

4.1 Computation of IPC Similarities

Similarities can be computed on different levels of the IPC hierarchy, i.e. on the class, subclass, group, or subgroup level. We computed the similarities on the subclass level in our work, which is the third level of the IPC hierarchy comprising 638 classes (in the current version IPC-2014.01). The IPC symbols on this level have four characters, starting with a letter for the section followed by a two-digit number for the class and a letter for the subclass (e.g. “A01B”). This four-character IPC symbol forms a common unit in patent retrieval and provides a good classification granularity. That is, the number of classes on this hierarchy level is ideal for the generation of IPC clouds, since they already contain a good amount of detailed information about the IPC class, but still retain a generality that provides an overview of potentially relevant IPC classes. However, the computation and mapping could also be performed on other levels of the IPC hierarchy.¹

To compute the similarities between the IPC symbols, we first build a vector space for the patent data. In our case, we used the 88,000 patents from the second repository of our database (see above). We created a vector for each of the 615 IPC symbols contained in that dataset², with the patents as dimensions of the vector space: If the considered IPC symbol is used to classify a patent, the corresponding dimension has a positive value; otherwise it is zero. Then, we compute the cosine similarity of each pair of IPC symbols to determine their relatedness in the patent data. That is, given two IPC symbols x and y , we first calculate the vectors V_x and V_y and subsequently compute their similarity with the formula

$$\text{sim}(V_x, V_y) = \frac{V_x \cdot V_y}{|V_x| \cdot |V_y|}. \quad (1)$$

¹In the following, we will also use the term *IPC symbol* when we refer to the shortened four-character version of the IPC symbol for the sake of simplicity.

²23 of the 638 available IPC symbols were not used in the dataset.

The cosine similarity is an efficient measure for sparse vectors, which is useful in our case, as each IPC symbol is associated with only a small fraction of the patents. This results in a small number of non-zero dimensions per vector compared to the total number of dimensions in the vector space, and hence in sparse vectors.

4.2 Dimensionality Reduction of IPC Space

In the second step, we map the IPC symbols onto a 2D plane required for the visualization. The goal of this step is to find a 2D representation that approximates the similarity matrix. That is, IPC symbols that are frequently co-used in the patent data are ideally placed close to each other, while those that never appear together are placed far apart.

Our implementation uses t-SNE [22] as mapping technique. We first normalize the similarity matrix to get a probability distribution P , where p_{ij} represents the similarity between IPC symbol i and IPC symbol j . The t-SNE algorithm aims to find positions $x_1, \dots, x_n \in \mathbb{R}^2$ which minimize the Kullback-Leibler divergence between two distributions P and Q :

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (2)$$

where q_{ij} is defined as:

$$q_{ij} = \frac{(1 + \|x_i - x_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|x_k - x_l\|^2)^{-1}} \quad (3)$$

representing the similarity between point x_i and x_j .

For the maximum number of iterations, we use the default parameter of 1000 [22].

5. IPC CLOUD VISUALIZATIONS

The 2D mapping of the IPC space provides the basis for the creation of IPC clouds. In particular, we developed two different types of IPC clouds that we call *map view* and *darts view* and that will be detailed in the following. While the map view provides a global overview on the IPC space, the darts view puts selected IPC symbols in the focus and supports the visual identification of IPC symbols that are related to the selected ones. Both views follow the “visual information seeking mantra” [20] by giving an overview first, then allowing to zoom and filter, and finally showing details on demand.

5.1 Map View

The map view is basically a normalized and rescaled depiction of the 2D representation we get after the dimensionality reduction. Additionally, the font sizes reflect the frequencies with which the IPC symbols are used.

If we would directly visualize the previously computed 2D representation of the IPC space, we would get many overlaps resulting from the fact that the text labels (i.e., the IPC symbols) have a non-zero width and height. As dimensionality reduction techniques typically map the data to an arbitrary Cartesian coordinate system, we first normalize and rescale the mapping. By doing so, we transform the mapping into a coordinate system appropriate for visualization, while we retain the spatial distribution. In our case, a scaling factor

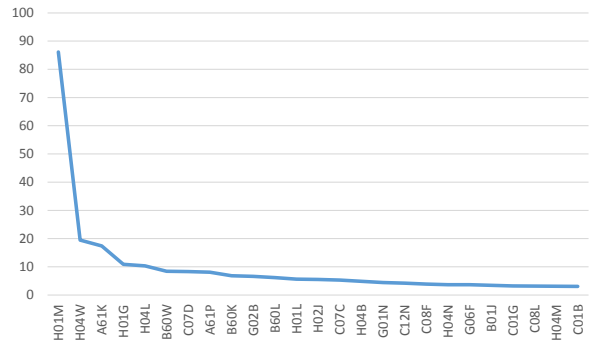


Figure 2: The distribution of the IPC usage frequencies roughly follows a power law, as illustrated for the 25 most often used IPC symbols in the 88.000 patent records that were analyzed (in thousands).

of 25,000 resulted in a good overview and only few overlaps of the text labels.

After the layout has been computed, the IPC symbols are placed at the determined positions on the screen, as shown in Figure 1(a). The font size of each IPC symbol correlates with the number of associated patents, i.e., IPC symbols with a large font size are used more often in the patent data than those with a small font size. We use a logarithmic scaling for the font sizes, as the frequencies of the IPC symbols roughly follow a power law distribution (cp. Figure 2) and we do not want to overemphasize certain IPC symbols. The resulting map view shows the whole IPC space, with the IPC symbols spatially arranged according to their relatedness and scaled in size according to their usage frequency.

In addition, we offer the user the option to remove even the few remaining overlaps, in case he or she wants to. We use the *push* variant of the Force-Scan Algorithm (FSA) [19] for this purpose, which preserves the general layout and, in particular, the relative distances of the nodes. The algorithm compares the label areas with each other and, if an overlap is detected, the label which is further to the upper left is fixed and all other labels are moved in the direction where the overlap is resolved the fastest.

Keeping the relative distances of the labels roughly stable is important, as they reflect the relatedness of the IPC symbols. This disqualifies many other algorithms for overlap removal that preserve the orthogonal ordering of the labels but not their relative distances [14]. A common drawback of the push variant of FSA is the increased size of the visualization, which is, however, not a problem in our case, as we usually expect only few label overlaps and as we added zooming and panning to the IPC clouds.

Panning and zooming are basic but important interaction techniques that enable the user to explore different parts of the map view in more detail. Furthermore, we added a minimap that always shows the whole IPC cloud and indicates which part of it is focused in the main view (Figure 1(c)). The minimap can also be used to change the focused area and to reset the zoom level. It therefore helps to avoid that the user gets lost in the IPC space.

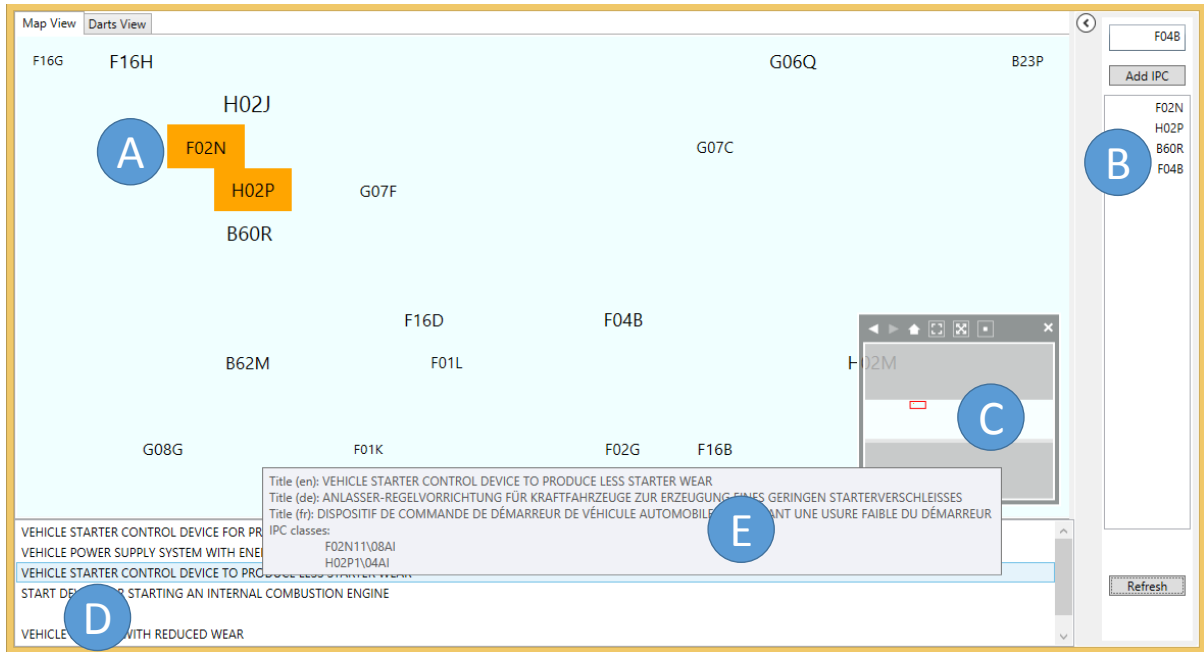


Figure 1: Map view of the IPC space where the user filtered four IPC symbols ①. These IPC symbols and their related ones are shown in the overview ②. The minimap ③ indicates which part of the IPC space is focused. The two highlighted IPC symbols have been selected by the user. The bottom part lists all patents that are associated with the selected IPC symbols ④. Further information about the patent, including all associated IPC symbols, can be displayed on demand ⑤.

Since users are typically interested in specific IPC symbols, they can filter the map view to show only a subset of IPC symbols and those that are co-used. This can be done by selecting any number of IPC symbols on the map and adding them to a whitelist displayed on the right of the visualization (Figure 1②). As it can be hard to spot specific IPC symbols on the map, the IPC symbols can alternatively be entered in a search field (equipped with an autocomplete feature). Once all IPC symbols of interest have been added and the filter is activated, IPC symbols that are not related to at least m of the whitelisted ones are removed from the visualization (with a variable m that is set to $m = 1$ by default).

If the user selects an IPC symbol in the visualization, the titles of patents associated with that symbol are listed beneath the main view (Figure 1④). If several IPC symbols are selected, only titles of patents associated with all of the symbols are listed (i.e. they are connected by a logical conjunction operator). More details on a patent, such as the whole list of associated IPC symbols and its titles in German and French, are shown in a tooltip when hovering over the patent's title in the list.

5.2 Darts View

The darts view provides another perspective on selected IPC symbols using the metaphor of a dartboard. In contrast to the map view, it does not provide a global overview on the IPC space but focuses on specific IPC symbols and their local context. IPC symbols selected in the map view or entered in the search field are placed in the center of the darts view (the bullseye), as they define what the user is inter-

ested in. Related IPC symbols are concentrically arranged around the bullseye in distances that reflect their relatedness to the selected IPC symbols: While IPC symbols close to the bullseye are strongly related, IPC symbols near the border have a weaker relation. Figure 3 shows an example where the IPC symbol “F02N” has been selected and hence forms the bullseye.

The darts view requires the definition of two key parameters: 1) a maximum number n of IPC symbols shown in the visualization, and 2) a threshold α defining the minimum similarity value a related IPC symbol must have to be shown in the visualization. Both parameters are interrelated and suitable values are dependent on the application context, such as the available screen space or the average font size of the labels. We had good experiences with an n of 10 to 20, as this number of IPC symbols can still be well perceived and cognitively processed. A good α value is more difficult to choose, as the similarity values are dependent on the considered patent data and IPC symbols. For our patent data, an α of 0.5 to 0.7 has led to good results in most cases. For instance, we used an α of 0.6 to generate the darts view shown in Figure 3. However, it could happen that for some IPC symbols no results are returned, as all similarity values are below the given threshold α .

Another option would be to dynamically choose an appropriate α based on the number of related IPC symbols that are returned. For instance, α could be dynamically changed in a way that there are always the n most related IPC symbols shown in the darts view. However, such an adaptive approach bears the risk that the user does not recognize the

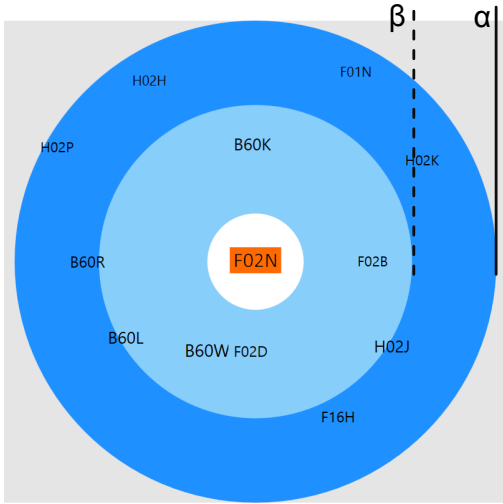


Figure 3: Darts view showing one selected IPC symbol in the bullseye and related IPC symbols concentrically arranged around it indicating their relatedness.

variable threshold when analyzing different darts views. It may also lead to a wrong impression, as the visualization might include IPC symbols that are only very distantly related to the selected ones in case of a low α .

After the related IPC symbols have been determined based on the parameters n and α , their positions on the dartboard are computed. Like the map view, the darts view makes use of the 2D representation we computed in Section 4, in that the related IPC symbols are located in the representation and their relative angle to the selected IPC symbol is determined. If multiple IPC symbols are selected, the average of the angles is taken. The related IPC symbols are then ordered by their angle. However, they are not drawn with their original angle on the dartboard but the angles are normalized in a way that they are forming a circle around the selected IPC symbol(s).

Apart from the angles, we also compute the distances of the IPC symbols in relation to the bullseye. We take the values that resulted from the similarity computation (cf. Section 4) and use a logarithmic scale to determine the final positions of the IPC labels. We decided for a logarithmic scale, as the similarities of the IPC symbols follow roughly a power law distribution again, i.e. the number of IPC symbols with a high similarity value is much lower than the number of IPC symbols with a lower similarity in nearly all cases. Finally, the IPC symbols are placed at the determined positions on the dartboard, while their font sizes indicate how often they are used in the patent data, like in the map view.

Note that there is no fixed value separating the inner from the outer circle of the dartboard by default. If we want to have such a value, we can simply define another threshold β for the inner circle (see Figure 3). This threshold β sets the borderline that separates IPC symbols in the inner circle from the outer. Likewise, we can add any number of additional circles to the darts view, each with its own threshold.

5.3 Example of Use

Let us assume we want to file a patent for a new technique to start combustion engines. The IPC symbol “F02N” is ideally suited to classify our invention, since it refers to the “starting of combustion engines” [13]. In the map view, we have already spotted said IPC symbol and noticed that the IPC symbol “H02P” is very close to it (as in Figure 1). It classifies patents that describe a “control or regulation of electric motors, generators, or dynamo-electric converters” [13].

We can therefore assume that several technologies for combustion engines are also used in electric motors. It seems to be a good idea to analyze the patents related to electrical engine starters, because there may already be a patent which is in conflict with our invention.

After switching to the darts view, we realize that there seem to be several other IPC symbols that are also strongly related to the IPC symbol we are interested in, leading us to further technologies and patents that might be of relevance and should be considered before filing our patent.

6. DISCUSSION OF SCALABILITY

Due to the massive number of patents that are digitally available nowadays, scalability is one of the main issues in any patent visualization approach. A key challenge in our approach lies with the 2D mapping of the IPC symbols. Dimensionality reduction methods are usually not stable, i.e. the algorithms may map data to very different locations on the 2D plane even if the data changes only slightly. Therefore, we do not recompute the 2D mapping with every change in the dataset but keep the mapping stable as long as it still reflects the IPC distances in a sufficient way. That is, stability has a higher priority than precision in this particular case, as the distances in the 2D representation only roughly indicate the relatedness of the IPC symbols anyway.

Besides the scalability of the visualization, the scalabilities of the data storage and data model are crucial in patent retrieval. The former is unproblematic in our approach, as new patent records can simply be added to the Elastic Search database. If new IPC symbols are added to the database, only those patent records need to be updated that are classified by these symbols, without the need to update any other patent records.

The data model is robust to an increasing amount of patents in the sense that the similarities of the IPC symbols do not need to be recomputed due to the usually large amount of patent records that are processed in the initial mapping. New patents will still be found if IPC symbols are selected in the visualization because the search for related patents uses the database without actually considering the data that has been used by the data model. This robustness entails two disadvantages: 1) it will be necessary to recompute the similarity matrix at some point, which will also require a remapping onto the 2D plane; 2) if a large number of patents will emerge in a specific field, such that the associated IPC symbols would get a lot more important, this approach would not be able to detect this shift in the IPC space. To represent new IPC symbols in the data model, it is necessary to recompute the similarity matrix as well as the 2D mapping of the IPC symbols.

	# of patents	# of IPC symbols
Data storage	+	+
Data model	Search: + Sim. accuracy: 0	-
Mapping	+	-

Table 1: Scalability of the data storage, data model, and mapping in relation with the number of patents and the number of IPC symbols.

Table 1 summarizes the discussed scalabilities of the various components of our approach. It indicates how well the data storage, data model, and mapping scale with an increasing amount of patents and IPC symbols after the initial computation of the data model.

7. CONCLUSION AND FUTURE WORK

We presented IPC clouds, an interactive visualization for the patent domain inspired by tag clouds that allows to explore the IPC space. In contrast to related work, IPC clouds do not make use of the predefined IPC hierarchy but are based on the actual co-use of IPC symbols in the patent data. They provide an overview of the IPC space and enable the user to ‘dive’ into it and find related IPC symbols that might be relevant in a specific retrieval context.

We presented two different types of IPC clouds: The *map view* arranges the IPC symbols globally on a 2D plane, while the *darts view* provides a local and focused layout for a selected subset of IPC symbols. It uses the metaphor of a dartboard with the selected IPC symbols in the bullseye and related symbols concentrically arranged around it. Although the visualizations look different, they are strongly related and can efficiently be created from the same 2D representation. Like in tag clouds, the font sizes of the IPC symbols are scaled according to their usage frequencies to emphasize IPC symbols that occur very frequently in the analyzed data. We added a simple search interface to the map view, using a whitelist of IPC symbols for filtering. Both visualizations are additionally equipped with several interaction techniques that support the exploration of the IPC space and allow to get more details about patents that are related to selected IPC symbols.

We are currently in the process of expanding our database to contain data for all patents indexed in Espacenet, which is more than 80 million [3]. Once these patents have been loaded into our database, we will investigate if there are distinguishable clusters or patterns of IPC symbols. We are also planning to extract concepts and components from the patent documents and visualize their relations in addition to the IPC space. Finally, we aim to extend and combine the map and darts view in a manner that they are integrated into one highly dynamic and interactive IPC cloud visualization.

8. ACKNOWLEDGMENTS

This work was partially supported by the EU funded project iPatDoc (grant no. 606163).

9. REFERENCES

- [1] Delphion Citation Link. <http://www.delphion.com/products/research/products-citelink>.

- [2] Elastic Search. <http://www.elasticsearch.org>.
- [3] EPO – Espacenet. <http://www.espacenet.com>.
- [4] EPO – European Publication Server. <https://data.epo.org/publication-server>.
- [5] EPO Worldwide Patent Statistical Database (PATSTAT). http://www.epo.org/searching/subscription/raw/product-14-24_de.html.
- [6] European Patent Register. <https://register.epo.org>.
- [7] IPC (International Patent Classification). <http://www.epo.org/searching/essentials/classification/ipc-reform.html>.
- [8] MongoDB. <http://www.mongodb.org/>.
- [9] Open Patent Services (OPS). <http://www.epo.org/searching/free/ops.html>.
- [10] PatAnalyse – Sample Patent Map. <http://www.patanalyse.com/samplemap.html>.
- [11] Patent iNSIGHT Pro. <http://www.patentinsightpro.com/>.
- [12] Thomson Innovation. <http://thomsonreuters.com/thomson-innovation>.
- [13] WIPO – World Intellectual Property Organization. <http://www.wipo.int>.
- [14] T. Dwyer, K. Marriott, and P. J. Stuckey. Fast node overlap removal. In *Proceedings of the 13th Int. Conf. on Graph Drawing*, GD’05, pages 153–164. Springer, 2006.
- [15] M. Giereth, S. Koch, M. Rotard, and T. Ertl. Web based visual exploration of patent information. In *Proceedings of the 11th Int. Conf. on Information Visualization*, IV ’07, pages 150–155. IEEE CS, 2007.
- [16] A. B. Jaffe and M. Trajtenberg. *Patents, Citations & Innovations: A Window on the Knowledge Economy*. MIT Press, revised edition, 2005.
- [17] D. O. Kutz. Examining the evolution and distribution of patent classifications. In *Proceedings of the 8th Int. Conf. on Information Visualisation*, IV ’04, pages 983–988. IEEE CS, 2004.
- [18] S. Lohmann, J. Ziegler, and L. Tetzlaff. Comparison of tag cloud layouts: Task-related performance and visual exploration. In *Proceedings of the 12th IFIP TC 13 Int. Conf. on Human-Computer Interaction, Part I, INTERACT ’09*, pages 392–404. Springer, 2009.
- [19] K. Misue, P. Eades, W. Lai, and K. Sugiyama. Layout adjustment and the mental map. *Journal of visual languages and computing*, 6(2):183–210, 1995.
- [20] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, VL ’96, pages 336–343. IEEE CS, 1996.
- [21] C. Sternitzke, A. Bartkowski, and R. Schramm. Visualizing patent statistics by means of social network analysis tools. *World Patent Information*, 30(2):115 – 131, 2008.
- [22] L. Van der Maaten and G. Hinton. Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.
- [23] F. B. Viégas and M. Wattenberg. Tag clouds and the case for vernacular visualization. *interactions*, 15(4):49–52, 2008.