

DOCUMENT MAPS FOR COMPETENCE MANAGEMENT

Timo Honkela

Helsinki University of Technology

Neural Networks Research Centre

Laboratory of Computer and Information Science

P.O.Box 5400, FI-02015 HUT (Espoo), Finland

timo.honkela@hut.fi

Raimo Nordfors

Gurusoft, Inc.

Tekniikantie 21 D

FI-02150 Espoo, Finland

raimo.nordfors@gurusoft.fi

Raimo Tuuli

Gurusoft, Inc.

Tekniikantie 21 D

FI-02150 Espoo, Finland

raimo.tuuli@gurusoft.fi

Abstract The WEBSOM is a method for analyzing and visualizing large document collections. In the WEBSOM method, the self-organizing map algorithm is used to automatically organize collections of documents onto a two-dimensional map to enable easy exploration and search of the collection. Map regions that are close to each other contain similar items. GS Textplorer is a software package that implements the WEBSOM method and provides a modular software architecture with separate processes for data input and preprocessing, document map creation and map use. GS Textplorer provides tools, e.g., for overall visualization and interactive exploration, classical keyword-based search and query-by-example approach, and analysis of a combination of qualitative and quantitative data. In this article, we will consider one successful application area of the document maps, i.e., competence management. The document map approach has proved to be useful, e.g., in internal and external recruiting and strategic competence analysis.

Keywords: Self-organizing map, human resource management, text mining.

1. Introduction

In the competence management process, an organization defines, captures and measures the existing competences and identifies development needs. In modern knowledge-intensive organizations, competence management is an important part of human resource management. A widely adopted approach for competence management is based on “traditional” knowledge management scheme with object-oriented, taxonomic and/or ontology-based knowledge creation, knowledge storage and retrieval, knowledge transfer, and knowledge utilization. Already Taylor (1911) stated that workers’ competencies should be handled by “classifying, tabulating, and reducing the knowledge to rules, laws and formulae.” Often the term competence is still nowadays used to denote a formalized set of roles and skill areas within an organization (consider, e.g., Vasconcelos, 2000). However, such classification frameworks are static in nature and they only relate to the current and past situation at best. Assessing and describing human competence is highly complex task due to the largely tacit and contextual nature of knowledge and skills (see, e.g., Nonaka & Takeuchi, 1995). Therefore, we claim that adaptive tools for data and text mining can be particularly useful in competence management enabling a more efficient process that can be based on a variety of heterogeneous sources of information. With suitable methods it is possible to analyze unstructured, semi-structured and structured information. In the following, we present how to apply adaptive systems in competence management. Our main focus is in the development of real-world applications that are based on text mining using the self-organizing map algorithm. We call the resulting systems document maps.

2. Document Map Methods and Tools

The WEBSOM is a method developed originally at Helsinki University of Technology for analyzing and visualizing large document collections (Honkela et al. 1996, Kaski et al. 1998, Lagus et al. 1999). In the WEBSOM method, the self-organizing map algorithm (Kohonen, 2001) is used to automatically organize collections of documents on a map to enable easy exploration and search of the collection (as related research consider, e.g, Lin et al., 1991, and Merkl, 1997). Map regions that are close to each other contain similar documents. The main objectives in developing the WEBSOM method has been to offer a method for exploring text collections that is different from the query-result approach, enabling the user to get an overall view to the document collection. In the following, we first describe the basic characteristics of the self-organizing map algorithm and then provide a description of the document map approach.

2.1 Self-Organizing Map

The self-organizing map (SOM) (Kohonen, 2001) is an artificial neural network model of the ordered “maps” in the cortex. In practical applications, it is widely used as a data mining and visualization method for complex data sets. Application areas include, for instance, image processing and speech recognition, process control, economical analysis, and diagnostics in industry and in medicine (see, e.g., Oja et al., 2003). The outcome in the SOM learning process is that ordered values for the prototype vectors emerge over the array. To demonstrate the basic properties of the SOM we can consider an example that is relevant in the human resource management area. Hofstede (1980) conducted a survey of more than 100,000 IBM employees around the world. He noticed differences between cultures and came up with theories to explain cultural variability. The basic dimensions to characterize the variability include: (1) uncertainty avoidance (how comfortable people feel towards ambiguity and risk), (2) power distance (extent to which members of an organization accept that power is distributed unequally throughout the organization), (3) masculinity-femininity (expected gender roles in a culture), and (4) individualism-collectivism (degree of individual or group orientation). Hofstede gained data from 50 countries. It is difficult to see the general relationships between the countries when the data is presented in a tabular form (see Table 1).

Table 1. An example of multidimensional data (Hofstede 1980).^{a,b}

<i>Country</i>	<i>Power distance</i>	<i>Individualism</i>	<i>Uncertainty avoidance</i>	<i>Masculinity</i>
Austria	11	55	70	79
Belgium	65	75	94	54
Brazil	69	38	76	49
...

The learning process of the self-organizing map is based on similarity comparisons in a continuous space, usually using Euclidean metric (see Kohonen, 2001 for details). The result is a system that associates similar inputs close to each other in the two-dimensional grid called the map. Thus, if two countries have similar values related to cultural aspects (Table 1) they tend to appear close to each other on the map. A map based on Hofstede’s (1980) data is presented in Fig. 1.

The advantage of the non-linear dimensionality reduction becomes even more clear when the number of dimensions is high, for instance 39, like in (Kaski and Kohonen, 1996). Similarly, the analysis and visualization of categorized competence data could be useful as such. However, in the following



Figure 1. Illustration of a self-organizing map based on data collected by Hofstede (1980). The clustering diagram shows the relationships between countries based on four cultural factors, namely, uncertainty avoidance, power distance, masculinity-femininity, and individualism-collectivism (see the text for more details). Relative distances in the original 4-dimensional space are illuminated by the shading: the darker an area on the map, the higher the distance (Ultsch, 1993). One can discern some rather obvious relationships such as the relative similarity between pairs Portugal and Greece, Sweden and Denmark, and the UK and the US.

we describe even more sophisticated approach based on the idea of creating self-organizing maps of text documents.

2.2 GS Textplorer

GS Textplorer enables automatic creation of maps of large number of documents (Honkela et al., 2003). Motivation for the GS Textplorer framework raises from the visual information retrieval paradigm, based on the WEBSOM method. The main features of GS Textplorer include modular architecture separating map creation and map use through a client-server solution, visualization of keyword search results on the map, and associative search of similar documents, i.e., query-by-example. The basic architecture is presented in Fig. 2.

The system supports knowledge discovery by, e.g., making it easy to find similar items to a relevant one, and providing an analyst an overall view on the text collection. One is able to see the relative frequency of various categories of data and their relationships. The user is also able to see development over time: which phenomena are growing, which are diminishing, which are coming closer to each other, and which are moving apart.

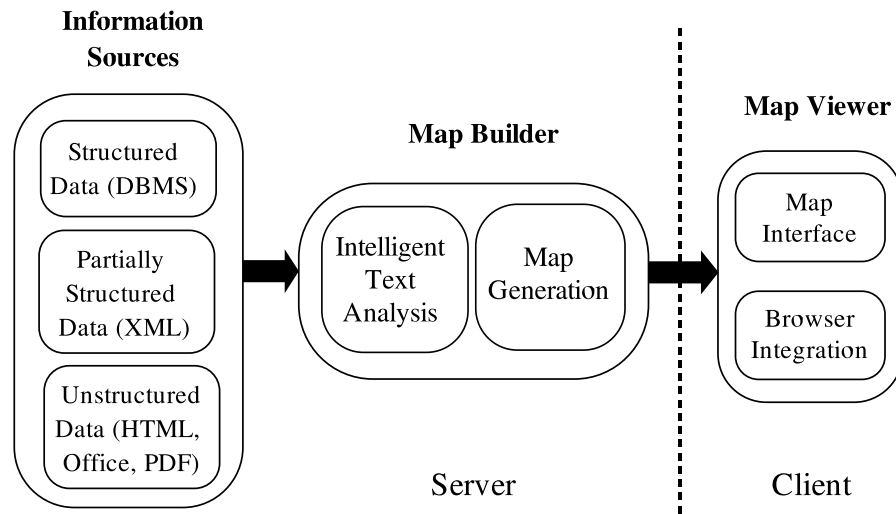


Figure 2. The general system architecture of GS Textplorer, presenting the basic phases of the automatic map building process. The map viewer provides a tool for exploration and search of the “competence space” based either on a stand-alone application or on an interface used through a web browser.

One of the most important potential benefits of document maps is the fact that no manual classification and categorization is needed. The method creates by itself the underlying conceptual space based on the document collection.

3. Document maps in competence management

In competence management traditional ontology-based, database-based or information retrieval solutions have some clear limitations as pointed out earlier in this article. For instance, looking for individual words or terms in an information retrieval systems is quite superficial approach. On the other hand, hand coding competence information is a huge effort. Such an approach relies on formalized knowledge and is thus both inefficient and rigid.

The knowledge creation process within document map approach is based on the idea that there is suitable data available that characterizes the competences accurately enough. In case of recruiting, the basic body of data consists of the CV of the applicants. Inside an organization there may be additional sources of information such as technical reports, project reports, performance and development discussions, and different assessment tools. The contents of texts written by a particular person appear to be a rather reliable source of information for the formation of his or her competence profile. Moreover, the inclusion of texts in the competence management system can be based on the self-service

principle. The dynamic nature of the analysis ensures that the features for determining the structure of the competence map follow dynamically the changes in the domain of the organization. No separate updating for the knowledge representation scheme, such as taxonomy, is needed: the competence landscape on a competence map is an emergent result of the overall qualities of the input material.

Next we illustrate the knowledge utilization process based on a competence map through some examples. For instance, one can analyze how well the competence profiles and the needs of the organization fit each other. There are several kinds of potential misfits described in the following. The study is based on consideration of a business context. Analogical considerations are valid for any type of organization.

- 1 There are persons who do not match any roles in business definition. This can lead either into education or reorientation of these persons or an update of the business plan to include a new area based on these skills. This latter update can be especially relevant if it can be estimated that the market opportunity would be considerable.
- 2 There are business roles for which there no people. This would lead either into recruitment of new people or education or reorientation of current personnel. The education option is especially relevant if there are persons whose have close interest for the role under consideration. In case neither of these options is realistic, one needs to consider if the role is really necessary to the business and if some modification for the business plan is needed.
- 3 If there is a long distance between the background of some persons and their interests, one can consider providing opportunities for practicing the interesting role or to provide some education towards that direction. This would most likely be beneficial considering the motivation of the persons. Fit with the business roles naturally needs to be considered.

Another specific context of use for competence mapping is in the case of a fusion of two companies or organizations. In the case of traditional competence classification systems, the effort needed for standardizing the systems and updating the data can be vast. When the competence management is based on free text documents, the reformation of the competence map for the new organization can be done automatically. In this article, we do not consider the combination of structured and unstructured data. However, the document map approach allows also the combination of different kinds of data.

For knowledge transfer the system can be used by several persons who can make task division regarding any information flow explicitly. A group of persons can share knowledge by annotating a shared map view. The map supports

straightforwardly information flow filtering. One can specify map areas that serve as trap: new items in the specified area can be redirected in any specified way.

One critical topic for the document map approach is the availability of data. If there is no relevant data available then the approach is not usable. However, it seems that in most knowledge intensive organizations this aspect is not relevant.

4. Conclusions

In the beginning of this article, we discussed the nature of “traditional” knowledge management. It is based on the idea that the world can be represented as a collection of objects and the relationships between the objects. This basic assumptions is present, e.g, in semantic networks, taxonomies, and ontologies. The network of knowledge is usually considered to have an objective nature, i.e., it is assumed that the users of the knowledge representation can apply it without any concern of problems that would be caused by individual points of view. The knowledge representations are commonly built manually which means that the specification and implementation of the competence classification system is laborious. Moreover, the classification system needs to be updated regularly because the domains of relevant competence are continuously changing because of developments in the domain and changes in the focus of the company or organization.

The approach for competence management based on document maps provides the following advantages.

- A database containing data relevant for competence management can be transformed with low costs into an illuminative competence landscape.
- It is possible to view a large collection of persons as a map that reveals otherwise hidden interrelationships.
- One can compare numerical and free-text data in the database.
- Candidates for a position (or a projects) can be chosen effectively through the map interface.
- It is easy to detect near matches: choosing candidates through the map interface without limiting oneself to those that exactly match the original query.
- The visual competence map provides efficient means for searching the candidates using query-by-example approach (i.e. looking for people "like Tom").

Acknowledgments

We are grateful to a large number of collaborators as well as current and former colleagues who we cannot list here.

References

- Hofstede, G.H. (1980). *Culture's Consequences: International Differences in Work-related Values*. Beverly Hills, California: Sage Publications.
- Honkela, T., Samuel, K., Lagus, K. & Kohonen, T. (1996). *Newsgroup Exploration with WEBSOM Method and Browsing Interface*. Report A32, Helsinki University of Technology, Laboratory of Computer and Information Science.
- Honkela, T., Honkela, J., Myyryläinen, M., & Tuulos, V. (2003). Adaptive and Visual Map Interface for Document Collections. *Adjunct Proceedings of Human-Computer Interaction International Conference*, pp. 119-120.
- Kaski, S. & Kohonen, T. (1996). Exploratory data analysis by the self-organizing map: Structures of welfare and poverty in the world. *Neural Networks in Financial Engineering*, Singapore: World Scientific, pp. 498-507.
- Kaski, S., Honkela, T., Lagus, K. & Kohonen, T. (1998). WEBSOM - self-organizing maps of document collections. *Neurocomputing*, 21:101-117.
- Kohonen, T. (2001). *Self-Organizing Maps*. Springer, 3rd edition.
- Lagus, K., Honkela, T., Kaski, S. & Kohonen, T. (1999). WEBSOM for Textual Data Mining. *Artificial Intelligence Review*, 13:345-364.
- Lin, X., Soergel, D., and Marchionini, G. (1991). A self-organizing semantic map for information retrieval. *Proceedings of 14th Annual International ACM/SIGIR Conference on Research & Development in Information Retrieval*, pp. 262-269.
- Lindgren, R. & Henfridsson, O. (2002). Using Competence Systems: Adoption Barriers and Design Suggestions. *Journal of Information and Knowledge Management*, 1(1), pp. 65-77
- Merkl, D. (1997). Lessons learned in text document classification. *Proceedings of WSOM'97, Workshop on Self-Organizing Maps*, Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland, pp. 316-32.
- Nonaka, I. & Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford: Oxford University Press.
- Oja, M., Kaski, S. & Kohonen, T. (2003). Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum, *Neural Computing Surveys*, 3: 1-156.
- Taylor, F.W. (1911). *The Principles of Scientific Management*. Available at <http://melbecon.unimelb.edu.au/het/taylor/sciman.htm> (downloaded February 16, 2004)
- Ultsch, A. (1993). Self-organizing neural networks for visualization and classification. Opitz, O., Lausen, B., and Klar, R., (eds.), *Information and Classification*, London, UK: Springer, pp. 307-313.
- Vasconcelos, J., Kimble, C., Gouveia, F. (2000). Corporate Knowledge Management using Competence Taxonomies. *Proceedings of Real World Taxonomies Conference*, Ark Group, London, September 2002.