DOMAIN SPECIFIC LANGUAGE IN TECHNICAL SOLUTION DOCUMENTS

Discussion of Two Approaches to Improve the Semi-automated Annotation

Helena X. Schmidt, Andreas Kohn and Udo Lindemann
Institute of Product Development, Technische Universität München, Boltzmannstraße 15, 85748, Garching, Germany

Keywords: Solution knowledge, Ontology-based knowledge system, Mechanical engineering, Domain specific language, Sublanguage, Annotation.

Abstract: The efficient search for existing solutions in mechanical engineering is a key-factor for successful product development. Ontology-based knowledge systems can support the semi-automated annotation of documents about existing solutions and enable the retrieval of those documents. However, the use of different wordings for similar products and a generally heterogeneous domain-specific language hinder the efficient annotation process. In this paper, two approaches to improve the semi-automatic annotation of documents by adding terms to the ontology are described. We evaluate the two approaches by analysing the industry sector-specific and company-specific languages used in documents in the field of mechanical engineering.

1 INTRODUCTION

In engineering, an important part of the product development process is the research for existing solutions. The discovery of existing solutions to a technical problem can shorten the product development process significantly. Certain barriers, such as unstructured data and the different use of language, hinder the access to existing solutions and increase the product developer's effort necessary for the search (Gaag et al., 2009); (Kohn et al., 2010).

These problems in solution research are addressed by the use case PROCESSUS which is part of the German research project THESEUS. Based on sales publications from the automation industry an ontology for representing knowledge about technical solutions has been developed. Solution documents from different industry sectors and different companies can be integrated into the ontology structure. An ontology-based prototype supports the semi-automated annotation of solution documents to integrate them into the ontology and the subsequent retrieval of relevant solution documents (Gaag et al., 2009).

With the prototype, the user can annotate solution documents semi-automatically. This can reduce the time needed for the annotation significantly. The text document is imported into the prototype. In combination with linguistic algorithms, such as word stemming and syntactic analysis, terms that already exist in the ontology as instances are recognized and suggested to the user. Therefore, the semi-automated annotation's success depends on the completeness of the instances contained in the ontology.

This paper focuses on improving the semi-automated annotation of solution documents by adding the appropriate instances for describing technical products within a solution to the ontology. As mentioned above, a factor for the identification of terms is the language used to describe them in solution documents. The language can differ from industry sector to industry sector and from company to company. The use of different terms and language in specific domains and its importance for knowledge management is a general challenge and has been addressed by scientists in different areas such as linguistics, information technology and engineering.

Therefore, this paper will first provide an overview of the research on language in specific domains. In the next step, challenges for the semi-automated annotation of technical solution documents are shown. Then, two approaches are explained and evaluated with an exemplary analysis of the language used in solution documents.

Schmidt H., Kohn A. and Lindemann U..
DOMAIN SPECIFIC LANGUAGE IN TECHNICAL SOLUTION DOCUMENTS - Discussion of Two Approaches to Improve the Semi-automated Annotation.
DOI: 10.5220/0003630301590166
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2011), pages 159-166
ISBN: 978-989-8425-80-5

2 LANGUAGE IN SPECIFIC DOMAINS AND ITS ROLE FOR KNOWLEDGE MANAGEMENT

In this section different approaches to the use of language in specific domains are presented. Then, the methods used for the analysis of language in this work are explained.

2.1 Research on Language in Specific Domains

Why is the examination of domain specific language important for knowledge management and – in particular – for ontology-based information retrieval systems? According to Thellefsen (2003), in systems for knowledge management the use of special language in documents is not considered sufficiently. He states that today, instead of adapting the system, the documents have to be adjusted to the system which leads to unefficient annotation and retrieval of knowledge.

To point out the importance of the context in which language is used several authors cite Wittgenstein. He introduced the term “language game” describing patterns in which a meaning of a word is explained by its use (Petras, 2006; Thellefsen, 2010).

This definition implies that one term can have different meanings in different contexts or as Petras states, ”there is no one-to-one mapping between a sign (term) and a concept (meaning)” (2006, p. 15). For knowledge management, this means that the same term can have different meanings for users with a different context.

Further semantic relations of terms play a role, such as synonymy (two terms with the same meaning) and hyponymy (one term is subordinate to another) (Vossen, 2003).

In linguistics there are different approaches to analyse the use of language in specific domains.

One approach is the study of sublanguages. It focuses on the syntatic and lexical characteristics of the language. According to Grisham and Kittredge (1986) a sublanguage is the language used by a community of speakers in a specific domain. Kittredge (2003) lists a number of characteristics of sublanguage such as a restricted lexicon and limited term co-occurrence structures. Losee and Haas (1995) establish statistic measures to evaluate the degree of specification of different sublanguages. They analyse abstracts from different scientific fields such as history and electric engineering and compare their occurrence and their meaning in specialized and general dictionaries. Another approach is the study of language for special purposes (LSP). The emphasis lies on the semantic characteristics (Petras, 2006). The characteristics analyzed are stylistic, such as the use of conditional sentences and the discursive line of the text (Evangelisti Allori, 2001).

In product development, linguistic analyses have been employed to facilitate the research for analogies from biology. Cheong et al. (2008) translate engineering terms into biological keywords to draw analogies from biology for solution research. The most promising biology terms are selected by their occurrence in biology dictionaries. Terms that occur very frequently are considered insignificant because they are too general, terms that occur very rarely because they are too specific.

Another research linked to product development was conducted by Bohm and Stone (2009). They propose an approach to map terms from a component database to terms of a component taxonomy. By comparing the similarity of the component’s naming terms and their function, they determine synonyms for the terms of the component taxonomy.

Summing up, most of the presented approaches focus on scientific texts. Bohm and Stone’s work is an exception, as their scope are documents provided by product developers for product developers. The research presented in this paper is focused on another type of documents: sales publications which describe technical solutions for the customer.

2.2 Methods for Analysing Language in Specific Domains

In linguistics, the term count is a parameter used for the analysis of language in text documents. The term count is the number of occurrences of a term in one or several documents. A theory states that the distribution of the term count in a number of documents approximately follows a Poisson distribution (Losee, 1995). The Poisson distribution is a statistic law for events that occur with a known average rate independently from each other. If the distribution of an event follows a Poisson distribution, its probability P that it occurs k times can be calculated by equation (1). l is the expected value, i.e. the arithmetic mean of occurrence (Härtter, 1974).

$$P(k) = \frac{l^k}{k!} * e^{-l} \tag{1}$$

Applied on term counts the occurrence k equals the term count. If term counts follow a Poisson distribution, the probability of a certain term count in a single document can be calculated (Losee, 1995).

3 CHALLENGES OF SEMI-AUTOMATED ANNOTATION

Solution documents for sales publications from the automation industry served as a basis for the development of the ontology (Gaag et al., 2009). Figure 1 shows the mapping of knowledge from the solution documents to the ontology structure. The boxes contain concepts which are connected by relations. The concrete instances are assigned to the concepts.

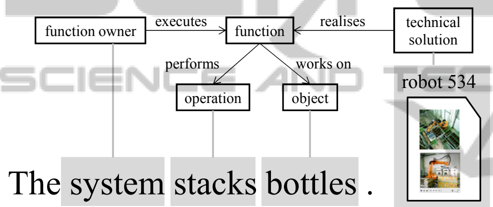

Figure 1: Mapping of solution knowledge to the ontology structure.

In common design theory, the research for solutions is usually based on the function the solution should perform (Ponn and Lindemann, 2008). Therefore, the function is set as the central concept in the ontology. A technical solution, e.g. the instance “robot 534”, realises a function. A function consists of the concepts operation and object. Figure 1 shows an exemplary sentence from a solution document. Here, the operation is “stack” and the object is “bottles”. The function is “stack bottles”. A function is conducted by a function owner. In this case the function owner is “system”.

As to the use of terms for function owners different concretization levels and the use of synonyms have been observed. As an example, we examine the function owner “system” from Figure 1. The term “system” does not describe the function owner’s characteristics. Instead of naming the function owner “system”, it can be concretised as ”robot”. It can be further concretized as “packing robot”, i.e. a term that concretises the function of the robot. There are more possibilities to concretize the term “robot”, for example a property of the robot can be described. A robot with articulations can be described as an “articulated robot”.

On the same concretization level a function owner can be described by a synonymous term. Instead of “system”, “equipment” or “machine” can be used. It depends on the context if the terms are synonymous. For example “machine” can be used as a synonym for “system” in the context of an automation system composed of several machine components. It seems less feasible to use it for systems with different system boundaries, i.e. bigger or smaller systems, for example - in the context of a valve system that is part of a machine itself. The synonyms can lead to other synonyms at a different concretization level. Instead of naming the function owner “packing robot” it can be referred to as “packing machine”, for example. In conclusion, to improve the semi-automated annotation the diverse terms used for function owners have to be added.

The two approaches described and discussed in this paper are the approach of noun extraction (section 4) and embedding classifications (section 5). The usefulness of the two approaches is evaluated for single companies and for several companies from the same industry sector. The applicability for different industry sectors is not regarded in this paper because the probability that different terms are used is higher. If an approach proves to be useful for several companies from one industry sector, its applicability for different industry sectors can be evaluated in a next step.

4 NOUN EXTRACTION

The following section details the procedure to identify function owners. The underlying assumption of this approach is that function owners found in the selection of solution documents also occur in unknown solution documents. The correctness of this assumption is evaluated in the second section.

4.1 Procedure to Identify Function Owners

Figure 2 shows the process of noun extraction. To start with, solution documents are selected. They have to cover different solutions the company offers to provide a representative selection of documents with a variety of function owners. As function owners in the ontology are nouns, the nouns from the solution documents are extracted. This step can be automated. The next step is the manual annotation performed by “experts”, i.e. users that are familiar with annotation. Then, the function owners

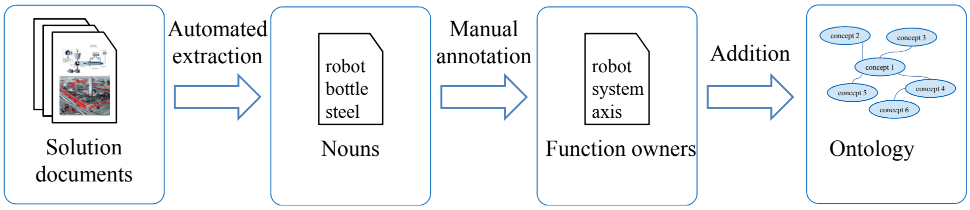

Figure 2: Noun extraction.

are added to the ontology.

4.2 Evaluation of the Approach

Is the assumption that function owners from a representative selection of solution documents also occur in unknown solution documents correct?

This can be evaluated on different levels. In this research, we analysed if

1. function owners extracted from one company occur in the solution documents from this company (company level).
2. function owners extracted from one company occur in the solution documents of a different company from the same industry sector (industry-sector level).

For the evaluation, a sample of German solution documents for sales publication of two companies from the automation industry was used. From company A 28 solution documents and from company B nine solution documents were used. Using a different number of solution documents allowed to observe if there is a correlation between the number of solution documents and the number of function owners that were extracted.

From the solution documents of company A 168 function owners, for the solution documents of company B 111 function owners were extracted using the procedure explained in section 4.1.

For the extraction of the nouns from the solution documents, the software tools *TreeTagger* developed by Schmid (2011) and *RapidMiner* from Rapid-I GmbH were used (Rapid-I GmbH, 2011) were used. Two scientific assistants annotated the function owners in the noun list.

The usefulness of the approach for the semi-automated annotation of unknown documents from the same company is evaluated in 4.2.1 (company level). Then, in 4.2.2, its usefulness for documents from the other company is evaluated (industry sector level).

4.2.1 Evaluation on the Company Level

For the evaluation, the term count of the extracted function owners in the solution documents is analysed (see section 2.2). It has to be differentiated between the term count of a function owner in all documents and its term count in the single documents. The distribution of the term count in the single documents shows how often a function owner occurs in how many documents. If a function owner occurs in a significant number of documents and is distributed regularly this is considered as an indication that it will occur in unknown solution documents as well. Both the significant number of documents and the regular distribution depend on the term count of the function owner in all documents.

In the next step, to further evaluate if an extracted function owner will occur in unknown solution documents, the real distribution of the term is compared to a Poisson distribution (see section 2.2).

Table 1 lists the nine function owners identified by noun extraction with the highest term counts in all documents for company A (Table 1). The function owners with the highest term counts in all documents are shown because a high term count is needed to analyse if the function owners occur in a significant number of documents and are regularly distributed. On the right side of Table 1 the distribution of the term count, i.e. the number of documents in which the function owner occurs with a certain term count is shown. To illustrate, the numbers are coloured. A dark colour means a high number of documents.

The most frequent function owner “robot” occurs 415 times in all documents (term count: 415). It occurs in all 28 documents at an average 14,82 times. “Robot” has a minimum term count of nine in one document and a maximum term count of 20 in four documents. Between minimum and maximum, every term count can be found in at least one document except the term count 19. Therefore it is concluded that “robot” has a relatively regular distribution over the 28 documents.

The second frequent function owner “gripper” occurs 66 times in all documents. It occurs in 24 documents which is considered a significant number

Table 1: Term counts of function owners from company A. Index: 1) Term count in all documents, 2) Number of documents that contain the function owner, 3) arithmetic mean of the term count in the single documents.

English term	1)	2)	3)		Term count single document																				
				number of documents	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
robot	415	28	14,82		0	0	0	0	0	0	0	0	0	1	1	2	4	3	4	1	5	1	1	0	4
gripper	66	24	2,36		4	7	3	7	4	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
industry robot	59	28	2,11		0	0	25	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
robot control	47	27	1,68		1	10	14	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
palletiser	45	15	1,61		13	6	2	2	1	1	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0
axis	33	13	1,18		15	8	2	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
robot cell	24	14	0,86		14	8	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
facility	21	15	0,75		13	9	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
roll conveyor	19	11	0,68		17	6	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

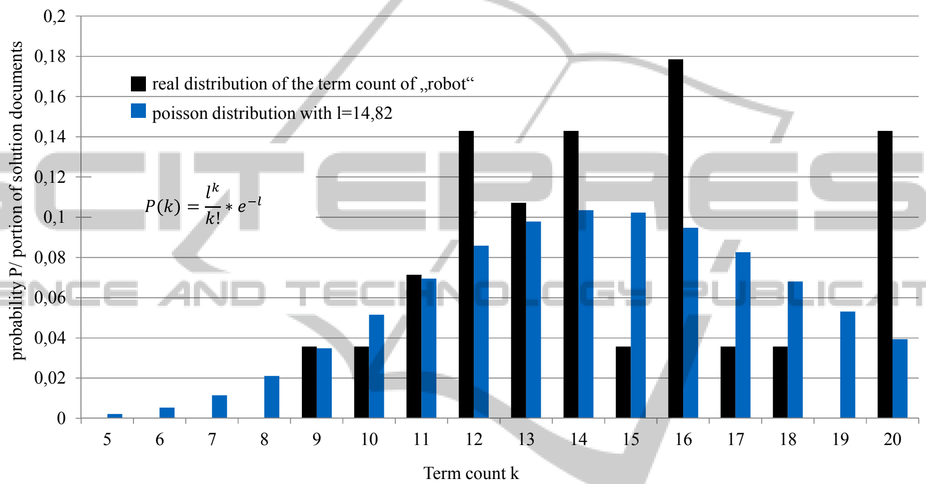

Figure 3: Real distribution compared to Poisson distribution.

of documents. The minimum term count is zero in four documents and the maximum term count is six in one document. In between minimum and maximum all term counts can be found in at least one document indicating a regular distribution.

This is the case for most of the next function owners, except “axis” and “roll conveyor”. Still, these two function owners have similar term counts in most documents.

As to company B, the term counts are lower than in the solution documents of company A due to the smaller number of documents. The function owners occur in more than one document but have a less regular distribution than function owners from company A.

In the next step, the distribution of a function owner’s term count is compared to a Poisson distribution to further evaluate if the conclusion to unknown solution documents is feasible.

As an example, the distribution of the function owner “robot” in the solution documents of company A is analysed. The arithmetic mean of its term count is 14,82. Consequently the expected value l of the adequate Poisson distribution is 14,82.

For the real distribution of the term counts the portion of documents with a certain term count is calculated. Figure 3 shows that the probability P and the portion of solution documents that have a certain term count are not equal. The real distribution of the term counts does not strictly follow a Poisson distribution. This result is in accordance with the results obtained by Losee’s analysis of scientific abstracts and the results of several other authors (Losee, 1995).

Even though a Poisson distribution was not found, both for company A and company B the majority of the most frequent function owners occur in a ”significant number of documents” and not only in one document. To sum up, it is inferred that extracted function owners that occur frequently in a sample of documents will also occur in unknown solution documents from the same company.

4.2.2 Evaluation on the Industry Sector Level

In accordance to the previous analysis for a single company, the term counts of the function owners from the two companies can be compared. Before comparing the distribution of the term counts in detail, it is examined how many function owners from company A equal function owners from company B. The result is that 14 function owners from company A were also extracted from company B. This amounts to 8 per cent of the function owners extracted from company A or 13 per cent extracted from company B. This low percentage of equal function owners shows that the noun extraction from documents from company A does not provide a significant number of function owners that occur in the solution documents from company B. Therefore, a further analysis of the term counts is considered unnecessary.

To analyse why the number of equal function owners in the documents of the two companies is so low, the function owners can be regarded with a focus on their meaning.

As an example, the function owner “robot” is examined. As explained in section 3, a function owner can have different concretization levels. In the solution documents from company A in addition to “robot” eight more concretely defined “robots” are used. In comparison, in the documents from company B two concretizations of “robot” are used. They are different from the concretizations of company A. The concretizations of company A and B are not synonymous. Examining the concretizations from company B more closely, the term count of “spot welding robot” is one and of “swivel-arm robot” is two. Both occur in one solution document respectively. This is an indication that they are very specific and that company B refers to most robots with the term “robot” without further concretizing them (term count of 72). Even though in documents from company A the term count of the general term “robot” is 415, concretizations such as “palletising robot” and “articulated robot” have a relatively high term count of 18 and 17. Thus, in documents from company A the function owner “robot” is concretized more often than in documents from company B.

This example shows how function owners are used differently by the two companies. For other function owners similar observations were made. Even though both companies are from the same industry sector and have similar products, they use different terms for their function owners. The use of language for function owners in solution documents from company A and B is company-specific. Therefore, the noun extraction of documents from one company does not provide a significant number of function owners relevant in documents from other companies.

5 EMBEDDING CLASSIFICATIONS

In this section the approach of embedding classifications is explained and evaluated for three different classifications (5.1). The results are compared in section 5.2.

For this approach terms from three different product classifications are added as function owners to the ontology. This has been described by Hepp (2005) for the *eCl@ss* classification.

5.1 Evaluation of the Approach

In this paper three classifications are regarded. The function owners from company A and B are compared to classes included in the classification. To justify the effort to embed classifications into the ontology, at least 40 % of the function owners should be included in a classification. For the comparison, the function owners obtained by noun extraction from company A and B were used. Hereafter, the three classifications are described and the results of the comparison are shown in Table 2.

5.1.1 VDMA e-Market

The VDMA e-Market is a platform provided by the VDMA, a German industry association with member companies from the capital goods industry. It contains approximately 200.000 product descriptions embedded in a classification (VDMA Verlag GmbH, 2011).

The *VDMA e-Market* classification is structured into eight industrial sectors. The sectors contain 1571 classes. The main product in a solution document is identified and assigned as an instance to a class (VDMA Verlag GmbH, 2011). As the objects are very specific product names, for the evaluation the function owners were compared to the classes in the *VDMA e-Market*.

5.1.2 eCl@ss

eCl@ss is a hierarchic system that classifies materials, products and services by standardized properties. It has been developed within a project funded by the German Ministry of Economy and Technology (eCl@ss e.V., 2011).

The *eCl@ss* classification contains classes divided into functional areas, main groups, groups and sub-groups. Products and services are classified on the subgroup level. Classes have properties with values, for example the length defined in mm. Key-words are assigned to classes. For the analysis, version 6.0 was used which contains 32592 classes and 51329 key words (eCl@ss e.V, 2011). For the evaluation, the classes were compared to the function owners.

5.1.3 UNSPSC®

UNSPSC® (*United Nations Standard Products and Services Code®*) is a standard for the classification of products and services developed within the United Nations Development Programme.

The classification includes 31498 classes numbered with a code (UNSPSC, 2011). For the evaluation, they were compared to the function owners.

5.2 Results

The results of the comparison are depicted in Table 2. The *UNSPSC®* classification includes 21 per cent of the function owners from company A and 33 per cent from company B. The other classifications contain less than 20 per cent of the function owners from company A and B. This is a relatively low percentage.

Table 2: Number of function owners included in classifications.

	Company A	Company B	both
VDMA e-Market	22 (13 %)	8 (7 %)	6 (43 %)
eCl@ss	16 (10 %)	13 (12 %)	4 (29 %)
UNSPSC®	35 (21 %)	37 (33 %)	12 (86 %)

In addition, the number of classes from the classifications is relatively high, especially for *eCl@ss* and *UNSPSC®*. Both classifications contain all types of products including food and plants.

The comparison of the function owners that company A and B have in common leads to a different result: Up to 86% of these 14 function owners occur in the classifications.

In conclusion, embedding the product classifications examined in this section does not improve the semi-automated annotation of the two companies significantly, because they contain less than 40 % of the examined function owners. The effort to include several thousands of classes into the ontology seems to be high in comparison. The function owners that are used by both companies are included with a significant percentage in the classifications. Still, as their number is relatively low this is no feasible improvement. With regard to the VDMA E-Market classification, the effort is lower which can justify embedding it, even though the improvement of the semi-automated annotation is slight.

6 DISCUSSION

The evaluation of the two approaches was performed with a small number of documents and companies. For the analysis of the term counts' distribution exact numbers for the "significant number of documents" and the "regular distribution" could not be stated. Consequently, the evaluation provides indications for the usefulness of the approaches for single companies and for different companies from the same industry sector. For general assumptions, a bigger document basis is needed which was not available at this point.

In addition to the evaluation of the two approaches, the conduction of the noun extraction provided insights about the strengths and weaknesses of noun extraction. Noun extraction can only extract nouns that consist of one word. Composite nouns such as "robot 75"and nouns containing hyphens such as "XY-robot" are not identified as one term. In German this poses fewer problems than in many other languages, because terms are often merged from several words. For example, "industry robot" is "Industrieroboter". During the manual annotation of the noun list, the annotation of the two scientific assistants differed on several function owners. A few function owners were overlooked by one assistant but annotated by the other. They disagreed if some nouns were function owners or not. In addition, a number of terms were annotated that are not used as function owners but as objects in the documents. An example is the term "work holding fixture". In common understanding this is a function owner, but in the analysed documents the term stands for an object which is assembled by a robot. Linguistic algorithms that distinguish between subjects and objects can be a solution to this problem. In most cases the function owner is the subject of the sentence whereas the object of a function is also "grammatically" an object.

On the other hand, a strong point of the noun extraction was the completeness of the list of function owners extracted. In previous research solution documents were manually annotated by several persons (Kohn et al., 2010). With the function owners identified by noun extraction of the manually annotated function owners up to 80 per cent of the manually annotated function owners are covered.

7 CONCLUSIONS AND OUTLOOK

In this work two approaches to improve the semi-automated annotation of solution documents in mechanical engineering were described and evaluated exemplarily. The first approach, the noun extraction, is promising if it is used to improve the semi-automated annotation of documents from the same company. For the annotation of documents from other companies from the same industry sector the results are not satisfying. This is due to company-specific use of language to describe function owners. The results for the approach of embedding existing classifications are less promising. The three regarded classifications contained a relatively low number of function owners.

This work discloses a number of starting points for future research. The noun extraction can be improved by applying linguistic algorithms to identify terms composed of several words and to distinguish between subjects and objects. For the embedding of classifications, other classifications can be regarded. As to the nature of function owners, the different specification levels could be further examined. In addition, synonyms can be added to the ontology.

ACKNOWLEDGEMENTS

Part of this work has been funded by the German Federal Ministry of Economy and Technology (BMWi) through THESEUS.

REFERENCES

Bohm, M. R. and Stone, R. B. (2009). A Natural Language to Component Term Methodology: Towards a Form based Concept Generation Tool. *ASME IDETC 2009*. San Diego, USA.

Cheong, H.; Shu, L.; Stone, R. B. (2008). Translating Terms of the Functional Basis into Biologically Meaningful Keywords. In *ASME IDETC,08.* Brooklyn, USA.

eCl@ss e.V.(2011). *eCl@ss – Das führende Klassifikationssystem*. Retrieved May 5, 2011 from http://www.eclass.de.

Evangelisti Allori, P. (2001). Conceptual and genre-specific constraints: How different disciplines select their discoursal features. In: Mayer, F. (ed.): Language for special purposes: perspectives for the new millenium. (pp. 70-79). Tübingen: *Gunter Narr Verlag*.

Gaag, A., Kohn, A. and Lindemann, U. (2009). Function-based Solution Retrieval and Semantic Search in Mechanical Engineering. In *ICED'09*. Stanford, California, USA.

Grisham, R., Kittrege, R. (1986). Analysing language in restricted domains. *Hillsdale: Lawrence Erlbaum*.

Härtter, E. (1974). Wahrscheinlichkeitsrechnung für Wirtschafts- und Naturwissenschaftler. *Göttingen: Vandenhoeck und Ruprecht*.

Hepp, M. (2005). A Methodology for Deriving OWL Ontologies from Products and Services Categorization Standards. In *ECIS 2005*. Regensburg, Germany.

Kittredge, R. I. (2003). Sublanguages and controlled languages. In Mitkov, R. (Ed.), The Oxford handbook of computational linguistics. (pp. 430-447). Oxford: *Oxford University Press*.

Kohn, A., Peter, G. and Lindemann, U. (2010). The Challenge of Automatically Annotating Solution Documents. In *KEOD'10*. Valencia, Spain.

Losee, R. and Haas, S. (1995). Sublanguage Terms: Dictionaries, Usage and Automatic Classification. *Journal of the American Society for Information Science*, 519-529.

Petras, V. (2006). Translating Dialects in Search: Mapping between Specialized Languages of Discourse and Documentary Languages. PhD thesis, *University of California*, Berkeley.

Ponn, J., and Lindemann, U. (2008). Konzeptentwicklung und Gestaltung technischer Produkte. (1rst ed.). *Berlin: Springer*.

Rapid-I GmbH (2011). *RapidMiner*. Retrieved January 13, 2011 from http://rapid-i.com.

Schmid, H. *TreeTagger* (2011). Retrieved January 20, 2011 from http://www.ims.uni-stuttgart.de

Thellefsen, M. (2003). The role of special language in relation to knowledge organization. *Proceedings of the American Society for Information Science and Technology*, 206-212.

Thellefsen, M. (2010). Knowledge Organization, Concepts, Signs: A Semeiotic Framework. PhD Thesis, *Royal School of Information and Library Science*, Denmark.

UNSPSC (2011). *UNSPSC®*. Retrieved May 5, 2011 from http://www.unspsc.org/

Vossen, P. (2003). Ontologies. In: Mitkov, R. (Ed.): The Oxford Handbook of Computational Linguistics. (pp.464-482). *Oxford: Oxford University Press*.

VDMA Verlag GmbH (2011). *Über den VDMA E-Market*. Retrieved May 5, 2011 from http://www.vdma-e-market.com/de/ueberEmarket.