

# Term Analysis – Improving the Quality of Learning and Application Documents in Engineering Design

S. Weiss, J. Jänsch, H. Birkhofer

*Conceptual homogeneity is one determinant of the quality of text documents. A concept remains the same if the words used (termini) change [1, 2]. In other words, termini can vary while the concept retains the same meaning. Human beings are able to handle concepts and termini because of their semantic network, which is able to connect termini to the actual context and thus identify the adequate meaning of the termini. Problems could arise when humans have to learn new content and correspondingly new concepts. Since the content is basically imparted by text via particular termini, it is a challenge to establish the right concept from the text with the termini. A term might be known, but have a different meaning [3, 4]. Therefore, it is very important to build up the correct understanding of concepts within a text. This is only possible when concepts are explained by the right termini, within an adequate context, and above all, homogeneously. So, when setting up or using text documents for teaching or application, it is essential to provide concept homogeneity.*

*Understandably, the quality of documents is, ceteris paribus, reciprocally proportional to variations of termini. Therefore, an analysis of variations of termini could form a basis for specific improvement of conceptual homogeneity.*

*Consequently, an exposition of variations of termini as control and improvement parameters is carried out in this investigation. This paper describes the functionality and the profit of a tool called TermAnalysis.*

*It also outlines the margins, typeface and other vital specifications necessary for authors preparing camera-ready papers for submission to the 5th International Conference on Advanced Engineering Design. The aim of this paper is to ensure that all readers are clear as to the uniformity required by the organizing committee and to ensure that readers' papers will be accepted as camera-ready for the conference.*

*TermAnalysis is a software tool developed within the pinngate project [5] by the authors of the paper at the department of product development and machine elements at Darmstadt (pmd) University of Technology. This tool is able to analyze arbitrarily and electronically represented text documents concerning the variation of termini. The similarity of termini is identified by using the Levensthein distance [6]. Identified variations are clustered and presented to the user of the tool. The number of variations provides the basis for identifying potentials of improvement with regard to conceptual homogeneity.*

*The use of TermAnalysis leads to the discovery of variations of termini and so generates awareness of this problem. Homogenization improves the document quality and reduces the uncontrolled growth of the concepts. This has a positive effect for the reader/learner and his/her comprehension of content [7]. By analyzing documents by various authors, a surprisingly high number of variations per document have been revealed. The investigations have indentified three main scenarios which are fully described in this paper.*

*Keywords: learning documents, product development knowledge, concepts.*

## 1 Introduction

The necessity for TermAnalysis in product development documents is given by the problem of termini in this knowledge domain [14], [15]. The problem of term non-homogeneity is also given in other knowledge domains, but in product development many knowledge domains come together and should work together. For this reason, product development uses termini from other domains with a new or changed meaning. Because of the non-homogeneity of termini in documents concerning product development, learning and teaching problems ensue. But the problem of terminology is not only an issue in education; it is also an obstacle to introducing product development knowledge in industry and other knowledge domains.

## 2 The pinngate-approach

Pinngate stands for product and process innovation gate, and is a current project of the department of product development and machine elements (pmd) at Darmstadt University of Technology. Pinngate is a teaching, learning and application environment. The main aim is to support dif-

ferent users with high quality information. The content is saved in a central database. The level of content is separated by the level of application by the so-called navigator. The navigator intervenes between these two levels and the user via a front-end (see Fig. 1) [10] [11] [12].

Based on this general concept, pinngate contains a number of tools that provide a range of supports for different users. One factor that this paper focuses on is quality, and one aspect of quality is the homogeneity of the termini.

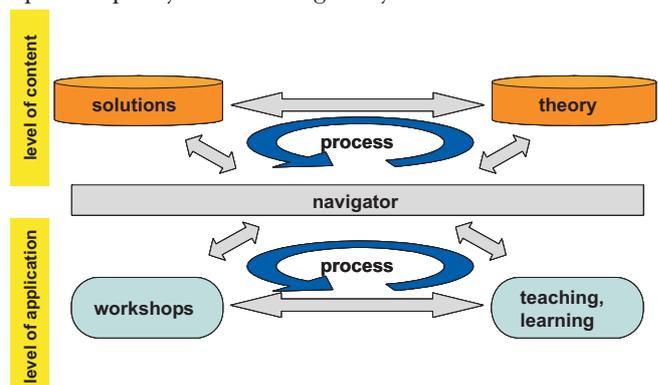


Fig. 1: The pinngate approach

### 3 Homogeneity of termini

Through the disposal of concepts a human being is able to think, learn and solve problems. The understanding of termini influences thinking, learning and problem-solving. A terminus is the name of a concept. A concept is more or less dependent on the individual and the situation. But how can termini and concepts be learned and taught? Generally, there are many rules for defining categories. According to the property theory, concepts are defined by accentuated properties. So, a particular object can be called a bird if it has two wings, feathers and a beak. If one of these things is missing, the object is not perceived as a bird. An object is categorized by comparing it to a prototype (representative example).

Ever since a study by Clark Hull (1920), a concept is understood as a category that has a certain system of classification. Accordingly, the learning of concepts consists of learning definitions and relevant properties. This method of learning concepts is based on the following assumptions:

- Each category is defined by a small number of relevant properties; the learner has to learn the relevant properties.
- An object only belongs to a certain category if it has the relevant properties.
- Within a certain level of abstraction, the individual categories are distinctly separated. An object cannot belong in two categories.
- The single properties do not differ according to their relevance. They all have the same relevance. [13]

Eleonor Rosch states that concepts can be systematized by natural conditions according to prototypes and best examples (ideal scenarios). A prototype is a representative example on a cognitive level that is generated from all the examples that have been observed. In this way an example is generated that best presents a concept. With additional rules the prototype can be specified and a certain degree of digression is possible. An example belongs to a concept if it fits to the ideal of the concept within a certain scope.

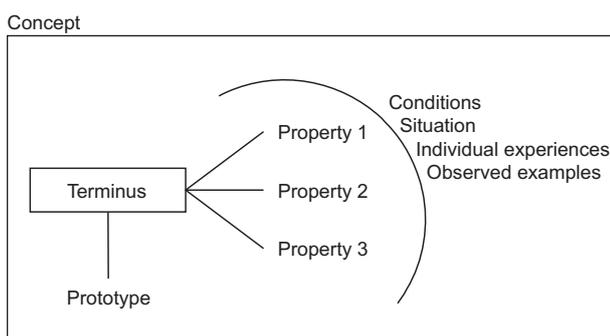


Fig. 2: Terminus vs. concept

Thus, the understanding of a concept depends strongly on the experiences made with the concept, the situation, problems and conditions. Looking at different knowledge domains, one and the same terminus can belong to different concepts and have different meanings (see Fig. 2).

For this reason, it is necessary to teach concepts adequately in the relevant knowledge domain under realistic conditions, situations, problems, etc. Further, it is absolutely

necessary always to use the same terminus for one and the same concept. If one uses different termini to describe the same concept, the learner starts to look for differences in the properties and tries to set up a second category or concept. This leads to confusion and misunderstanding [16]. Therefore, it is absolutely necessary to retain a high homogeneity of concepts within documents of learning material. The quality of learning documents becomes strongly diminished if homogeneity of termini is not considered. Homogeneity of termini is not the only prerequisite for good documents. Termini and their corresponding concepts must also be properly introduced with a sensible amount of adequate examples and instructions.

### 4 Term analysis

The quality  $Q$  of documents may be understood as a function of different parameters  $x_i$  influencing the quality. One of these parameters is the homogeneity of termini  $B$ . This can be written as  $Q = f(x_1, x_2, \dots, x_i)$ . Now we set each parameter different from  $x_i = B$  constant:  $x_y = \bar{x}_y \forall y = 1..i - 1$ . So, homogeneity is *ceteris paribus* the only determinant in the following argumentation.

Pinngate deals with various documents, which can generally be understood as objects. Each content item is represented in a modular way. A central modularization approach gives us a strategy for dealing with content divided into smaller modular constituents. However, content can be modularized or unmodularized. Modularized content can always be transformed into unmodularized content through reconstruction according to the modularization approach. Moreover, content can be newly created or it can already be present in the system's database. Within the argumentation of homogeneous termini, it is necessary to check new content before it is saved in the database. It must also be possible to check already existing content for homogeneity of termini. Thus, the task is defined: Create a draft that fulfills the following requirements:

- Identification of variations of termini
- Structuring of identified variations
- Applicable to both modularized and unmodularized content
- Applicable to new content
- Applicable to already existing content
- Applicable to any electronic text
- Compatible with pinngate

Based on these requirements, an approach should be developed that analyzes documents and identifies variations of termini, so that new documents are homogeneous from the very beginning. Existing documents can also be analyzed and improved using this approach. The basic strategy of TermAnalysis is summarized in Fig. 3.

The draft requires a textual document to be analyzed in five steps.

#### Step 1: Creating the Potential Term List

An algorithm analyzes the input file and identifies all the termini. Rules have to be defined on how to process symbols and special characters, such as & or -. On this occasion it is important to isolate each term exactly one time so that there will

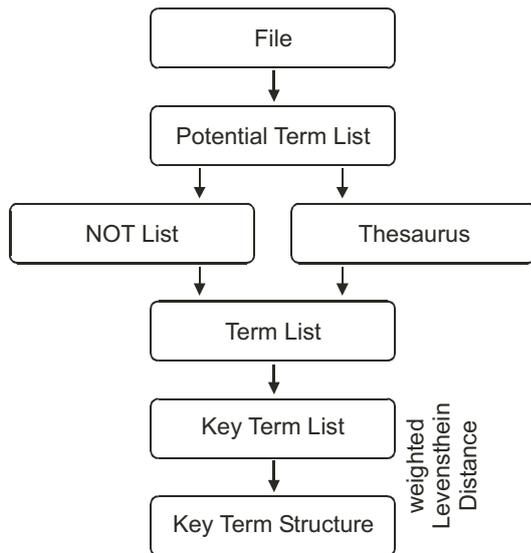


Fig. 3: Basic Strategy of TermAnalysis

be no redundancies later in the process. This is important because the whole process runs much faster based on a smaller Potential Term List. One has to think ahead to the comparisons of identified termini. Thus, the result is the so-called Potential Term List. This is a list in which each term used in the original file is represented exactly one time.

### Step 2: Applying the NOT List and the Thesaurus

TermAnalysis uses two additional techniques to deal with the Potential Term List: the NOT List and a thesaurus. The NOT List is a list that contains a collection of termini which are not technically termini, such as articles and prepositions, as well as termini that should not be processed further on. This is a predefined filter mechanism to remove irrelevant termini from the Potential Term List. The NOT List is pre-configured and can be modified by the user. The application of the NOT List reduces the Potential Term List. The thesaurus maps termini and their synonyms. So, different termini can be treated as one and the Potential Term List is reduced again. The thesaurus is predefined but can also be modified by the user.

### Step 3: The Term List

The Term List is the list representing all the remaining termini. It is the basis for identifying the variations of termini. The smaller the Term List is, the faster the variations can be determined. The Term List gives the user an overview of all important words used in the original file. It is recommended to sort the Term List alphabetically and evaluate it manually to get an impression of the words that are used. This allows one to draw a first conclusion about the quality of the original documents.

### Step 4: The Key Term List

The Key Term List contains very important termini that should be at the center of the subsequent analysis. The Key Term List is the basis for the algorithm to be applied in the next step. The main idea is to gain speed. Thus, the algorithm does not compare each term from the Term List with each other. Rather, it compares each term from the Term List with each Key Term of the Key Term List. The Key Term List is

predefined but can be – indeed, must be – modified by the user. The definition of the Key Term List sets the focus on the real important termini that the user wants to analyze. Moreover, the Key Term List can be created automatically. This is done by an algorithm identifying the most frequent strings or substrings.

### Step 5: Creating the Key Term Structure

The creation of the Key Term Structure is the final step on the way from the original file to the variation of termini. Each Term in the Term List will be compared with each Key Term of the Key Term List. This is done by calculating the weighted Levensthein Distance.

“Although there are many models for similarity among words, the most generally accepted in text retrieval is the Levensthein distance, or simply, edit distance. The edit distance between two strings is the minimum number of character insertions, deletions, and replacements needed to make them equal.”[6]

The algorithm used in TermAnalysis uses the weighted Levensthein Distance, i.e. different weights are considered concerning insertions, deletions and replacements. The result of the comparison is a tree based on Key Terms and their variations obtained from the Term List.

These five steps result in different information concerning homogeneity of termini. The following section gives an overview of the results that can be achieved by applying Term Analysis to documents.

## 5 Results

A first result is the impression gained by manually analyzing the alphabetically sorted Term List. Mostly, it can be determined that special termini have been used very often in different phenotypes. Moreover, it is possible to identify first Key Terms manually. This impression is a first sign of how consistent your choice of words really is.

However it is more impartial to derive a statistical overview of the results, so that it is transparent how often each term has really been used and which variations of it have been built. These results are a good platform for discussing the authors’ original document. It is also a good basis for improving the document. Especially in the case of learning documents, variations of termini should be minimized, because such variations may confuse the students.

These statistics can be generated for the whole document or chapter by chapter. So, one gains an overview of peaks of variations depending on the chapter that one looks at. This may indicate Key Terms, too, because each chapter deals with specific problems and the used termini depend on the problem. Thus, a peak of variations identified within a specific chapter allows one to conclude that the Key Terms are critical: either there is no clear definition of the concept or the authors have used it sloppily.

Especially in the context of learning documents, it is very important to use concepts well. Each key concept has to be used very carefully because this has an impact on the students. The students have no chance to determine whether one term is synonymous with another or not, and therefore, cannot distinguish different termini representing the same concept. Moreover, it could happen that the student recognizes differ-

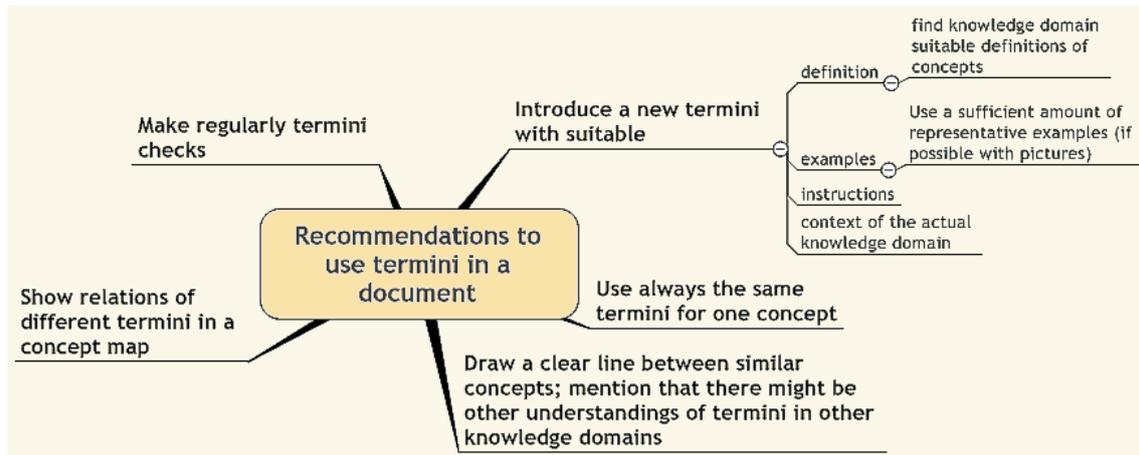


Fig. 4: Recommendations for introducing and using termini in documents

ent individual concepts represented by different termini and memorizes them.

This can be dealt with very easily within the pinngate project. Within pinngate, content is saved and modularly processed. Thus, the definition of each concept is given modularly, too [9]. Each document is also modularly represented modularized, so it is easy to determine two important positions: first, the position of the first occurrence of any concept and of associated termini, and second, the position of the modular definition of the concept. So, it must be stated that the modular definition of any concept has to occur at an earlier position than its variations of termini. Then there is a good chance that the students will not be confused, even if there are still variations of termini present.

TermAnalysis supports the author in analyzing his/her work and minimizing variations of termini. It facilitates the writing and reworking of documents. It helps to identify inconsistencies of termini and their definitions. With these advantages, TermAnalysis contributes to the improvement of product development knowledge and supports transfer of knowledge to students, industry and other domains.

## 6 Example and consequences

An example of a tool to support the consistency of terms within a text, after TermAnalysis has identified different termini for one concept is a concept map. Fig. 4 shows a concept map (also called mind map) that gives recommendations on how to integrate termini, especially technical termini, in a document. Concept mapping makes it possible to emphasize relevant properties of concepts and to distinguish them from each other.

## 7 Conclusions

To use TermAnalysis properly, the document has to be available in electronic form. It is sensible to have a well structured file system with various documents. This paper shows that the Levenstein algorithm is suitable to check the termini consistency of documents. The checking speed of the TermAnalysis tool runs up to seconds for 100 words (depending on the hardware). The tool only examines the consistency,

not the quality of the content. TermAnalysis is very useful for authors of learning and teaching documents. In most cases, the authors of such documents are experts, and therefore, very familiar with concepts and termini. But they also make use of „internal“ (insider) termini or use different termini for one concept without realizing it. Thus, TermAnalysis can also be seen as a tool of knowledge engineering that helps to externalize experts' knowledge properly.

## References

- [1] Specht, G.: *Einführung in die Betriebswirtschaftslehre*. Stuttgart: Poeschel, 1990, p. 14.
- [2] Seiffert, H.: *Einführung in die Wissenschaftstheorie I*. München: Beck, 1969, p. 37, 41.
- [3] Strube, G.: *Wörterbuch der Kognitionswissenschaft*, Stuttgart: Klett-Cotta, 1996, p. 58.
- [4] Seel, N., M.: *Psychologie des Lernens*, München: Ernst Reinhardt Verlag, 2003, p. 166.
- [5] www.pinngate.de
- [6] Baeza-Yates, R., Ribeiro-Neto, B.: *Modern Information Retrieval*. ACM Press, 1999, p. 105.
- [7] Anderson, J.: *Kognitive Psychologie*, Heidelberg, Berlin, Oxford: Spektrum Akademischer Verlag, 1996.
- [8] Weiß, S.: *Konzept und Umsetzung eines Navigators für Wissen in der Produktentwicklung*. Düsseldorf: VDI-Verlag 2006.
- [9] Birkhofer, H., Weiß, S., Berger, B.: Modularized Learning Documents for Product Development in Education at the Darmstadt University of Technology. In: Proceedings of DESIGN 2004, Dubrovnik, 2004, p. 599–604.
- [10] Weiß, S., Berger, B., Jänsch, J., Birkhofer, H.: COSECO (Context-Sensitive-Connector) – A Logical Component For a User- and Usage-Related Dosage of Knowledge. In: Proceedings of ICED 03, Stockholm, 2003.
- [11] Weiß, S., Berger, B., Birkhofer, H.: Topology of Modular Knowledge Structures in Product Development. In: Proceedings of DESIGN 2004, Dubrovnik, 2004.
- [12] Jänsch, J., Sauer, T., Walter, S., Birkhofer, H.: User-Suitable Transfer Of Design Methods. In: Proceedings of ICED 2003, Stockholm, 2003.

- [13] Mietzel, G.: *Pädagogische Psychologie des Lernens und Lehrens*, Göttingen: Hogrefe, 2003.
- [14] Hansen, F.: *Konstruktionssystematik*, Berlin: VEB Verlag Technik, 1965.
- [15] Birkhofer, H., Kloberdanz, H., Berger, B., Sauer, T.: Cleaning Up Design Methods – Describing Methods Completely and Standardised. In: Marjanovic, D. (Hg.). DESIGN 2002. Vol. 1. Faculty of Mechanical Engineering and Naval, Zagreb, The Design Society, Glasgow: Dubrovnik, Croatia. p. 17–22.
- [16] Jänsch, J.: Akzeptanz und Anwendung von Konstruktionsmethoden im industriellen Einsatz – Analyse und Empfehlungen aus kognitionswissenschaftlicher Sicht, Dissertation, Fortschritt-Berichte VDI, Reihe 1, Nr. 396, Technische Universität Darmstadt, VDI-Verlag, Düsseldorf 2007.

---

Dipl.-Wirtsch.-Ing. Sascha Weiß  
phone: + 49 (0) 6151 – 16 2666  
fax: + 49 (0) 6151 – 16 3355  
e-mail: weiss@pmd-tu-darmstadt.de

Dipl.-Wirtsch.-Ing. Judith Jänsch  
phone: + 49 (0) 6151 – 16 3055  
fax: + 49 (0) 6151 – 16 3355  
e-mail: jaensch@pmd-tu-darmstadt.de

Dept. of Product Development and Machine Elements  
(pmd)

Darmstadt, University of Technology  
Magdalenenstraße 4  
64289 Darmstadt, Germany