DERIVATION OF DESCRIPTION FEATURES FOR ENGINEERING CHANGE REQUEST BY AID OF LATENT DIRICHLET ALLOCATION

M. Riesener, C. Dölle, M. Mendl-Heinisch, G. Schuh and A. Keuper ⊠

RWTH Aachen University, Germany

⊠ a.keuper@wzl.rwth-aachen.de

Abstract

Complex products and shorter development cycles lead to an increasing number of engineering changes. In order to be able to process these changes more effectively and efficiently, this paper develops a description model as a first step towards a data driven approach of processing engineering change requests. The description model is systematically derived from literature using text mining and natural language processing techniques. An example of the application is given by an automated classification based on similarity calculations between new and historic engineering change requests.

Keywords: engineering change, product development, data mining, latent dirichlet allocation, automated classification

1. Introduction

Technological progress and the increasing speed of innovation are resulting in significantly shorter development cycles (Schuh et al., 2016a). Companies must offer products at competitive prices within the shortest possible time and with fewer resources (Schuh et al., 2016b). Technological innovations are accompanied by changing customer requirements. Nowadays customers expect highly developed, individual yet affordable products (Schuh and Riesener, 2018). To meet these expectations, many products are evolving into highly complex systems that merge the latest technologies from the fields of mechanics, electronics and software (Schuh and Riesener, 2018). These trends favour the occurrence of technical changes, which are an inseparable part of every development process and have a significant influence on the efficiency and effectiveness of development processes, as they account for around 25% of development capacity (Potdar and Jonnalagedda, 2018; Langer et al., 2012). To achieve competitive advantages it is crucial to manage engineering changes as efficient as possible (Wilberg et al., 2015). Currently, an average of 76% of all stakeholders involved in the development process are also involved in the handling of critical technical changes (Langer et al., 2012). The average processing time for a technical change is 15% of the whole project duration (Langer et al., 2012). Although the scope of research in recent years included the optimisation of processes related with technical changes, even more rigorous approaches are needed to exploit the identified saving potentials.

A suitable approach to make processes more efficient and more effective, is the systematic use of data and data analytics (Fan et al., 2014). Advances in data analytics and more powerful computers enable companies to generate value from the data recorded (Wickel, 2017). An application of data analytics in the field of technical changes has great potential for saving time and resources (Wickel et al., 2015).

Automated decisions or decisions supported by data analysis could reduce both, the number of stakeholders involved and the duration of the change process.

Therefore, the aim of the paper is to develop a description model of engineering change requests (ECR). With the description model ECRs can be systematically characterised and similarities can be identified which speeds up the change process since the same teams may be assigned to similar change requests. Furthermore, it is an important first step towards a fully automated evaluation process of ECRs. The application of the description model is shown exemplarily through automated classification of ECRs within the description model.

The remainder of the paper is structured as follows. After the introduction, section 1 continues with a description of the relevant terminology in this paper. Furthermore, a generic description of data mining techniques is given, to identify useful methods for application in the context of ECRs. Section 2 analyses related research of ECR description models and techniques to identify descriptive dimensions of ECRs. Furthermore, the research gaps addressed by this paper are derived. Section 3 provides an overview of the research method used in this paper to develop the description model and the procedure for automated classification of ECRs. Afterwards, the results of the introduced research method are presented in section 4. The last section gives a critical reflection and draws a conclusion.

Relevant terminology

In order to enclose the focus of the research paper, a short definition for the terms "engineering change", "engineering change request" and "data mining" is given and an overview of data mining techniques is presented.

Engineering Change:
There are multiple definitions of an engineering change (EC), one of the most cited is the definition of Jarratt et al. (2011). Therefore, it is regarded as generally accepted and will be used as the definition of EC in this paper:

> *"An engineering change is an alteration made to parts, drawings or software that have already been released during the product design process. The change can be of any size or type; the change can involve any number of people and take any length of time" (Jarratt et al., 2011)*

Engineering Change Request:
An engineering change request (ECR) is a request, created to inform about a derivation from target values (Feldhusen and Grote, 2013). It is created to support the decision process within the engineering change process. The three possible outcomes of an ECR are: Approval, dismissal or rework (Riesener et al., 2019). The purpose of an ECR is summarised and defined as:

> *"The engineering change request (ECR) is a request for a technical change in the state of a product model that has already been released for further development or production." (Riesener et al., 2019)*

Data Mining:
A contemporary definition of data mining was published by Schuh et al. (2019), where data mining is defined as:

> *"DM is a process aiming to generate knowledge from data and presenting findings comprehensively to the user, knowledge here means the discovery of new and non-trivial patterns, relations and trends in data useful to the user." (Schuh et al., 2019)*

Data mining actually originates from the Knowledge Discovery in Databases (KDD) according to Fayyad et al. (1996), which aims to identify new and useful patterns in data. While the KDD is rather scientifically and academically characterised, Shearer (2000) developed the Cross-Industry Standard Process for Data Mining (CRISP-DM), which focuses on the application of data analysis in the business environment. The steps of the iterative CRISP-DM cycle are the following: Business understanding, data understanding, data preparation, modelling, evaluation and deployment.

Text Mining:

The typical form of an ECR consists of fields for free text entry, therefore it is plausible to assume the format of input data for a data analysis of ECRs to be text-based. There is a specific subcategory of data mining, which is focused on text-based input data named "text mining". The translated definition from Müller and Lenz (2013) describes text mining as follows:

> *"Text Mining involves statistical or machine learning and analysis procedures to uncover and visualize latent structures and content in unstructured or weakly structured text data - such as TXT, PDF, Word, HTML or XML formats - in a given document collection." (Müller and Lenz, 2013)*

Natural Language Processing:

In order to use data mining algorithms on text-based input data, the text has to be transformed into a format understandable for computers. This is done by the so-called natural language processing (NLP):

> *"Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications." (Liddy, 2001)*

Possible tasks and applications for natural language processing are: Text categorisation, text filtering, text summary, spelling and grammar checking, automated translations or information extraction. In order to fulfil these tasks, there are several preparation steps for nearly every NLP-task (Figure 1).
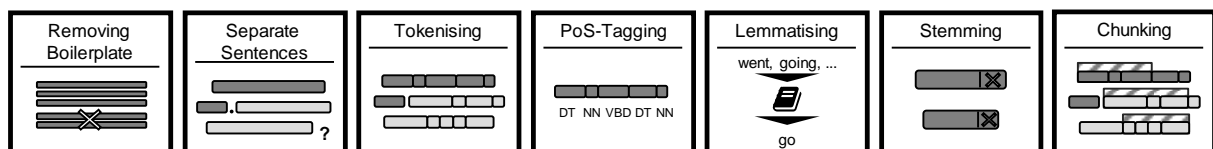

Figure 1. Steps of natural language processing (Müller and Lenz, 2013; Mustafi, 2016)

For detailed information of natural language processing steps it is referred to the work of Bird et al. (2009), Müller and Lenz (2013) or Mustafi (2016).

After defining the terms EC, ECR, data mining, text mining as well as NLP the next section of the paper will derive research gaps through analysis of related research. Afterwards, techniques of the introduced fields of text mining and NLP are used to develop a description model of ECRs and a method for automated classification of new ECRs.

2. Related research

The Analysis of related research must consider two different fields. On the one hand existing description models of ECs and ECRs have to be considered. On the other hand, it is necessary to explore data-based methods to find a suitable method for the identification of the relevant ECR characteristics and an approach for automated classification.

Prior research to identify characteristics of EC has been done by several authors. For example Eversheim et al. (1997) developed a description model for EC, which was used to identify different types of changes that occur within business networks. Their work investigated the processes of every network partner in regard to the type of EC. The aim was the creation of standard processes for the network for every type of EC to reduce time and cost of the change process within network cooperation. Aßmann (2000) focused on the optimisation of change processes inside a company. Three types of changes were defined and a description model to classify changes to one of these types was developed. Like in the work of Eversheim et al. (1997) processes for every type of ECR were defined. Wänström and Jonsson (2006) developed a framework based on a literature review to describe ECs and their influence on the material planning processes. Five perspectives were identified as relevant for this purpose: change, product, manufacturing, supply and demand. Within each perspective several features that influence the material planning were found. Belener (2008) developed a description model and a method to classify ECs by means of the concerned module, the propagated changes and the causes of the change, in order to give a prediction of the estimated required resources.

The classification has to be done manually with extensive examination of the EC. In the work of Hamraz (2013) a scheme that connects the function-behaviour structure with the change prediction method in order to predict change propagation is developed. Also, a comprehensive description model of EC causes was developed. More characteristics of ECs were not investigated in depth.

Even though there are many approaches that develop description models of ECR, all description models were developed with a specific purpose and therefore focus on a few aspects, instead of conducting a systematic approach to identify characteristics of ECs and ECRs holistically. In addition, Riesener et al. (2019) identified the need for a description model in the process of evaluating ECRs. The theory deficit concerning description models can be summarised as follows: A holistic description model considering all aspects of ECRs is not present. This deficit will be addressed by this paper with a systematic approach to identify all aspects of ECRs and developing a comprehensive description model.

After identifying the theory deficit concerning a holistic description model, the remaining of this section will focus on methods that can be used or adapted for the cause of this paper. A suitable approach to calculate similarities between ECRs is presented by Sharafi (2013). A transformation of the ECRs into the Bag-of-Words (BoW) vector format was performed in order to uses cosine similarity calculations and cluster algorithms on these vectors. The focus was to find patterns within the ECR, but the similarity calculation was not used for an automated classification approach. An automated evaluation of ECR is performed by Metha (2010), which calculates similarities between ECRs with a set of important attributes. The impact of a new ECR is calculated and scaled based on the most similar ECRs. Also, an automated evaluation was implemented, the classification of the attributes and the allocation of values to the attributes has to be done manually. Grieco et al. (2017) used a self-organising map on the Bag-of-Words vector representation of ECRs to cluster groups of similar content. The clusters divided the documents according to concerned part or assembly but the assignment of the clusters to the corresponding component needs to be done manually. Approaches that used NLP and data-based methods in the field of ECs are focused on finding patterns and predicting the impact of ECR. There was no approach identified that automatically classified ECR within a description model.

The analysis shows, that a method to systematically identify attributes of ECR does not exist in the field of ECs. By extending the search into other fields an approach of Rebentisch et al. (2018) was identified which presented a systematic approach to derive aspects of agile culture from literature. A latent dirichlet allocation (LDA) algorithm was used to discover latent topics in a collection of relevant literature. This systematic and objective method can be adapted for the cause of this paper, to derive characteristics of ECRs from literature and create a holistic description model. The theory deficit concerning data-based methods can be summarised as follows: The application of NLP-based similarity calculation for automated classification of ECRs and the application of LDA in order to systematically derive attributes of ECRs for the description model.

After analysing related research, the identified deficits will be addressed by developing a holistic description model of ECRs using LDA to derive attributes from literature. Within this description model ECRs can be automatically classified through NLP based similarity calculations between new and historic ECRs.

3. Research method

This section describes the methodology used to develop the ECR description model and the automated classification. As mentioned before, the overall aim of this paper is the support of decision making within the EC process through automated classification of ECR in a comprehensive description model. In order to identify the aspects and attributes of ECR the first step is to derive descriptions, characteristic and definitions of ECR from literature with the technique of a systematic literature review (Borrego et al., 2014). The resulting pieces of information are used as input for the topic modelling with the LDA. The LDA, first introduced by Blei et al. (2003), is a machine learning algorithm that has been widely used in text mining and other domains (Huang et al., 2018; Rebentisch et al., 2018; Wei and Croft, 2006). The generated topics are interpreted, and attributes are derived. After the consolidation of the derived attributes the description model can be developed.

Through a research project with a plant manufacturer the authors had access to 284 ECRs of that plant manufacturer. For the preparation of the automated classification 244 of these ECRs were manually

classified within the description model. In the last step for 40 of these ECRs an automated classification is conducted through similarity calculations. An overview of the six different steps of the methodology is presented in Figure 2 and further described in the following section.
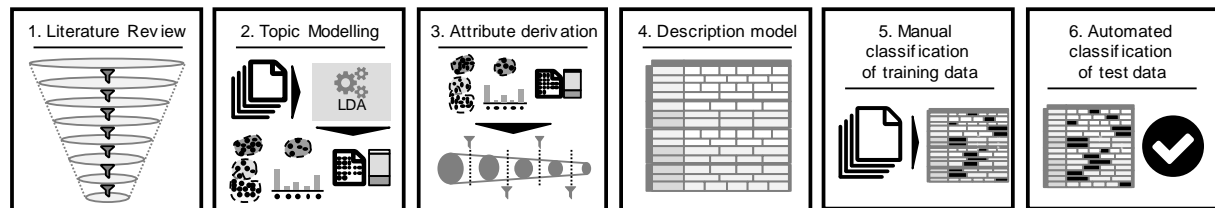

Figure 2. Methodology of this paper

Step 1 - Systematic literature review for the identification of relevant information
The systematic literature review is conducted to identify relevant information and provide it as input data for the LDA in the next step. It is conducted as described in Borrego et al. (2014) and adapted to the field of engineering design by Riesener et al. (2019) and Petersen (2017). A systematic literature review attempts to reduce subjectivity through the totality of the works and the associated perspectives (Borrego et al., 2014). In order to be able to identify as many relevant pieces of literature as possible, it is necessary to vary both the search terms and the databases. The identified documents need to be further analysed and filtered to narrow them down to the actual relevant results.

Step 2 - Topic modelling
In step 2 topics are generated based on the text extracts from the literature review. In order to analyse the text extracts with LDA, the text extracts have to be transferred to vectors in the BoW format (Salton and Buckley, 1988; Blei et al., 2003). Therefore, several NLP steps are necessary: lower casing all letters, part of speech (POS)-tagging and keeping only nouns, adjectives, verbs and adverbs, removing stop words, lemmatising and tokenising. Next, a dictionary is created with all remaining different words. The transformation of the text extracts to the BoW format creates vectors with the dimensionality equal to the number of entries in the dictionary. The value of the entries is between 0 and 1. The simplest variant of a BoW representation is to set the value of an entry to 1 if the document contains the corresponding word from the dictionary, if the word is not contained the entry is set to 0 (Roelleke and Wang, 2008). Other variants like tf-idf calculate the entries based on the frequency of the occurrence of the word within the document and the whole collection of documents (Roelleke and Wang, 2008).
The LDA assumes that documents (text extracts) are represented as a probability distribution of topics, where each topic is a probability distribution of words (Blei et al., 2003). The probability distributions are dirichlet distributions, which can be influenced by parameters α and η (Morik et al., 2015). LDA tries to iteratively find the words in the topics that are most likely to form the given documents with given probability distributions (Blei et al., 2003). To objectively compare the results of LDA generated topics, coherence values were introduced, representing the interpretability by humans (Röder et al., 2015). The highest coherence values for the documents in this paper were achieved with the parameters $\alpha = 0.5$ and $\eta = 0.01$, which are close to recommended values in literature (Wei and Croft, 2006).

Step 3 - Derivation of attributes
In the next step, attributes of ECRs are derived through the interpretation of the LDA topics. Besides these attributes the topics of the LDA can give an overview of the perspectives on ECRs. Build on these perspectives additional literature was analysed to find more attributes. Additional attributes can be derived through the analysis of the 284 ECRs from a plant manufacturer. Through removing the duplicates, consolidation and checking if the attribute is suitable to classify ECRs the list can be narrowed down. Through comparison of pairs the most relevant attributes can be identified.

Step 4 - Description model
After the most relevant attributes of ECRs are identified, the characteristic values (CV) of these attributes are investigated through analysis of literature and an analysis of the mentioned 284 ECRs. The result of this step is the developed description model.

Step 5 - Manual classification of training data sets

In the fifth step, the description model is validated by classifying 244 of the 284 ECR in it. It is checked whether the attributes and characteristic values derived from theory are present in the examples from actual ECRs of a plant manufacturer.

Step 6 - Automated classification of test data set

In the last step a concept is developed to automatically classify the remaining 40 ECRs within the description model by calculating similarities between the ECRs already classified and new ECRs. In order to check the classification accuracy, the ECRs are then manually classified and the match with the automated classification is calculated.

4. Results

In this section the results and interim results of the research method introduced in section 3 are shown.

Step 1 - Systematic literature review for the identification of relevant information

In the first step a systematic literature review is conducted as described in section 3. Table 1 summarises the procedure in which 194 relevant text extracts were determined.

Table 1. Steps and results of the systematic literature review

Step	Description	Result
1 - Synonyms for the search terms	In the first step for the search string "Description Engineering Change Request" synonyms are searched for "Description" and "Engineering Change Request". All combinations of the found synonyms are used as search terms in the literature databases of Google Scholar, Springer Link, Web of Science and Scopus	43 different search terms
2 - Export of search results	With the 43 search terms there are 363108 search results in total. By using the filters on the websites of the literature databases (e.g. field, year, …) and by removing duplicates a list of 15457 potentially relevant documents was exported	15457 documents
3 - Filter stop words	The next filter looks for stop words in the titles of the documents. The list of stop words contains words to further sort out documents from different domains like biology, medicine, politics or law	6494 documents
4 - Reading filter	The titles of the documents are checked manually	799 documents
5 - 2nd reading filter	The abstracts of the documents are checked manually	501 documents
6 - Full text analysis	The full text of the remaining documents is analysed and relevant parts are extracted	194 relevant text extracts

Step 2 - Topic modelling

The next step analyses the text extracts with the topic modelling algorithm LDA. The topics are represented through the 30 most relevant words within each topics. Overall, there are 31 interpretable topics, some of them occurred more than once, which signalises a higher importance (Table 2).

Table 2. Subjects of the LDA generated topics

Change environment	Description of ECR	Aspects of engineering change management
Company specific factors (1)	Reason of change (6)	Engineering change management (2)
External company environment (1)	Extend of change (3)	Realisation of change (1)
Product development & design (2)	Object of change (4)	Positive effects of change (1)
Manufacturing (1)	Initiator of change (1)	Negative effects of change (1)
Life-cycle (1)	Priority (3)	Elements of ECR forms (2)
	Involved Stakeholders (1)	

An example for the interpretation of the words of a topic is given in the following.

Example: A topic labelled as "Reason of change" consisted of the following words: Initiate; Customer; Problem; Development; Requirement; Source; Company; Require; Innovation; Emergent; Reason; Technology; Modification; Market; Improvement; Error; Management; Research; Field; Legislation.

DESIGN ORGANISATION AND MANAGEMENT

Step 3 - Derivation of attributes

As described in Section 3, attributes from three sources are aggregated into a longlist of 168 attributes and then reduced by various filters. The results of this procedure is visualised in Figure 3. In the end the 14 most relevant attributes are identified and included in the description model.
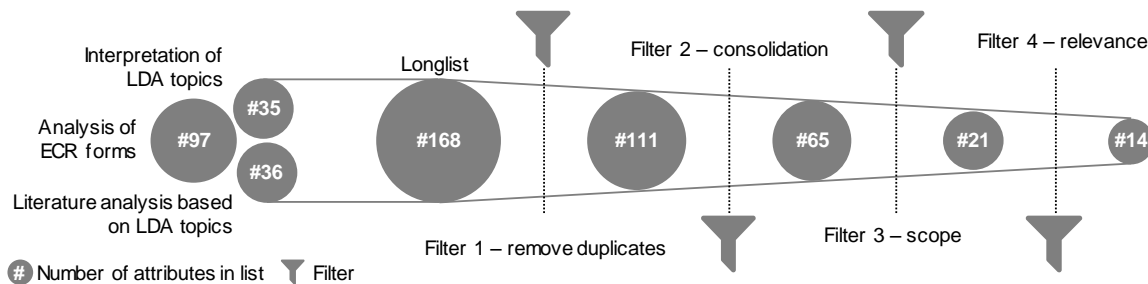

Figure 3. Procedure for finding the most relevant attributes

Step 4 - Description model

In accordance to the 14 prior identified attributes, characteristic values have been identified through analysis of literature and ECRs. The result is the description model depicted in Figure 4. The attributes have been grouped into the groups "Object of Change", "Properties [of the Object of Change]", "Influencing Factors [on the EC]" and "Effects [of the EC]". A distinction is made between two types of attributes. The descriptive attributes can be clearly identified using the engineering change request, whereas the evaluating attributes need to be assigned with context or expert knowledge.

	Attribute	Characteristic value					
Object of change	Concerned object	Document	Component	Process	Module	Equipment	Product
	Concerned object specification	Physical Specification	Functionality	Software	Procedures	Information	
	Extend of change*	Document	Component	Process	Module	Equipment	Product
	Volume of change	Single position		Group		All	
Properties	Phase of life-cycle	Development	Prototyping / Series start-up	Series production		Use phase	
	Architecture	Modular		Partly integrated		Integrated	
	Complexity	Low		Medium		High	
Influencing factors	Root of change	Customer requirement	Law / Certification	Safety	Cost	Technology	Productivity
	Reason of change	Current situation faulty	Optimization of the current situation	Target requirements changed		Incorrect target requirements	
	Urgency	Change to next release		Important, short-term change		Immediate change	
	Concerned Stakeholder	R&D	R&D and other departments		Involvement of the management	Involvement of external stakeholders	
Effect	Effect on cost	Decreasing	Slightly decreasing		Slightly increasing	Increasing	
	Effect on time schedules	Low		Medium		High	
	Effect on the product	None	Necessary correction	Fulfillment of individual customer requirements	Improvement of functionality	Quality improvement	

▢ Descriptive attributes

▨ Evaluating attributes (expert knowledge required)

*The CVs of concerned object and extend of change are the same. The diffrence is, that concerned object identifies the objects from where the change originates, while extend of the change describes objects that are impacted by the change of the object. *Example: If the dimension of a component (=object) is changed, most likely the drawing document will be changed too (=extend of change)*

Figure 4. Resulting description model

Step 5 - Manual classification of training data sets

In the next step the 244 of 284 available ECRs are classified manually within the description model. The classification had to be done without context knowledge of the company and therefore the evaluating attributes and "Phase of life-cycle" (due to missing project schedules) were not considered. Since the assignment of characteristic values to the ECRs was possible in all cases, the results show that the developed description model is suitable to reflect the characteristics of the ECRs.

Step 6 - Automated classification of test data set

After the classification of the 244 ECR training data sets is known, similarity values are calculated between a new ECR and all ECRs that are assigned to a characteristic value. The new ECR is assigned

to the characteristic value that contains the ECRs with the greatest similarity to the new ECR. The similarity is calculated based on the input fields of the ECR forms: Reason for change, component/assembly, description of current situation, benefit/goal of the change, proposed realisation. The first two fields are dropdown selections and the last three fields are free text fields. For every attribute a different field may have bigger influence, therefore a weighting factor is included and specified for every attribute. The process is summarised in Figure 5.
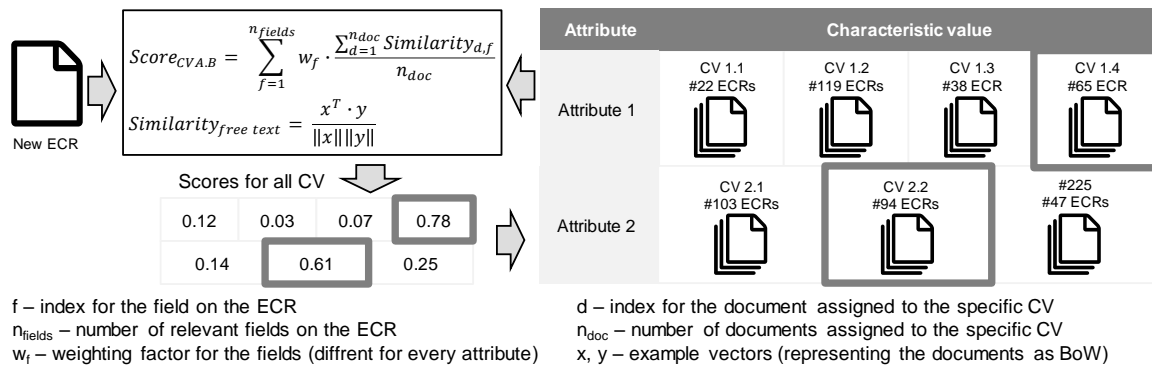

$$Score_{CV A.B} = \sum_{f=1}^{n_{fields}} w_f \cdot \frac{\sum_{d=1}^{n_{doc}} Similarity_{d,f}}{n_{doc}}$$

$$Similarity_{free\ text} = \frac{x^T \cdot y}{\|x\|\|y\|}$$

f – index for the field on the ECR
n_{fields} – number of relevant fields on the ECR
w_f – weighting factor for the fields (diffrent for every attribute)

d – index for the document assigned to the specific CV
n_{doc} – number of documents assigned to the specific CV
x, y – example vectors (representing the documents as BoW)

Figure 5. Method for automated classification of new ECRs

For the best results, the weighting factors of the five input fields of the ECR (Figure 6) have to be optimised. The values of the weighting factors and the amount of correctly, automatically assigned characteristic values are shown in Figure 6. An overall accuracy of 79% is accomplished.

		Weighting factors for the fields of the ECR					Automated classification (n=40)	
		Reason	Component/ Assembly	Description	Benefit/ Goal	Proposed Realisation	# of correct assignments	accuracy
Attributes	Concerned object	0,059	0,139	0,329	0,263	0,252	30	75,0%
	Concerned object spec.	0,077	0,181	0,044	0,779	0,157	35	87,5%
	Extend of the change	0,183	0,562	0,040	0,026	0,392	28	70,0%
	Volume of the change	0,012	0,107	0,069	0,630	0,011	25	62,5%
	Root of the change	0,334	0,435	0,209	0,014	0,152	35	87,5%
	Reason of the Change	0,238	0,316	0,411	0,330	0	36	90,0%

Figure 6. Weighting factors and correct classifications

Reading example Figure 6: In order to classify the ECR regarding the reason of change, the ECR field of description is most relevant (weighting factor of 0.411), whereas the proposed realisation is completely irrelevant (weighting factor of 0). With these weighting factors it is possible to correctly classify 36 of 40 ECR.

5. Conclusion

The description model developed within this paper shows the feasibility of LDA to identify attributes for a description model. The presented approach is suitable to identify relevant themes in a collection of documents and minimises the influence of the personal opinion of the creator of a description model. Since the effort to prepare the documents is currently very high, there is a need for research to improve the usability in terms of selection and preparation of the input documents for the systematic derivation of description models using LDA. A more extensive database than 194 text extracts used in this paper is recommended, to generate more meaningful and comprehensible topics.

In addition, this work presents an application example for the description model by using it to classify ECRs automatically. This contributes significantly to the examination of the feasibility of a completely automated processing of ECRs where the systematic description of ECRs is an essential first step. The current classification accuracy of 79% is expected to increase with a larger number of data sets. Besides, if company experts classify training data, a more accurate and therefore better classification can be achieved. When having access to a significantly larger database, the method of automated classification

can possibly be further improved using advanced algorithms like an artificial neural network (ANN). An ANN is much more adaptable in this respect and greater precision is to be expected.

In the future, the utilisation of the developed description model will be within a framework for data based processing of ECRs. With data mining methods patterns between the assigned characteristic values in the description model and resources spent for the realisation of the EC can be identified. This will support the process of predicting the cost when a new ECR arises. It is also possible to automatically assign the processing of ECRs to teams based on similarities with prior processed ECRs of that team.

Currently, the description model can be used as a foundation for the development of new ECR templates that support easier classification of ECRs and therefore easier data driven processing.

In general, this paper also showed that the use of text mining and natural language processing in the field of engineering design processes can be beneficial and should be considered when optimising engineering design processes in the future.

Acknowledgement

References

Aßmann, A. (2000). *Gestaltung von Änderungsprozessen in der Produktentwicklung*, *Produktentwicklung*, Vol. 38 No. 2. unveränderte Auflage, Utz, Herbert, München.

Belener, P.M. (2008), *Technisches Änderungsmanagement modularer Produkte und Prozesse*, Shaker Verlag, Aachen.

Bird, S., Klein, E. and Loper, E. (2009), *Natural Language Processing with Python*, Safari Books Online, O'Reilly Media Inc, Sebastopol.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent Dirichlet Allocation", *J. Mach. Learn. Res.*, Vol. 3, pp. 993-1022.

Borrego, M., Foster, M.J. and Froyd, J.E. (2014), "Systematic Literature Reviews in Engineering Education and Other Developing Interdisciplinary Fields", *Journal of Engineering Education*, Vol. 103 No. 1, pp. 45-76. https://doi.org/10.1002/jee.20038

Eversheim, W., Warnke, L. and Schröder, T. (1997), "Änderungsmanagement in Entwicklungskooperationen", *VDI-Z*, Vol. 139 No. 3, pp. 60-63.

Fan, J., Han, F. and Liu, H. (2014), "Challenges of Big Data Analysis", *National science review*, Vol. 1 No. 2, pp. 293-314, https://doi.org/10.1093/nsr/nwt032

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996), "From Data Mining to Knowledge Discovery in Databases", *AI Magazine*, Vol. 17 No. 3, p. 37. https://doi.org/10.1609/aimag.v17i3.1230

Feldhusen, J. and Grote, K.-H. (2013), *Pahl/Beitz Konstruktionslehre: Methoden und Anwendung erfolgreicher Produktentwicklung, 8., vollständig überarbeitete Auflage*, Springer Vieweg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-29569-0

Grieco, A., Pacella, M. and Blaco, M. (2017), "On the Application of Text Clustering in Engineering Change Process", *Procedia CIRP*, Vol. 62, pp. 187-192. https://doi.org/10.1016/j.procir.2016.06.019

Hamraz, B. (2013), *Engineering change modelling using a function-behaviour-structure scheme*, Apollo - University of Cambridge Repository, 2013. https://doi.org/10.17863/CAM.14061

Huang, A.H. et al. (2018), "Analyst Information Discovery and Interpretation Roles. A Topic Modeling Approach", *Management Science*, Vol. 64 No. 6, pp. 2833-2855. https://doi.org/10.1287/mnsc.2017.2751

Jarratt, T.A.W. et al. (2011), "Engineering change. An overview and perspective on the literature", *Research in Engineering Design*, Vol. 22 No. 2, pp. 103-124. https://doi.org/10.1007/s00163-010-0097-y

Langer, S. et al. (2012), Änderungsmanagement-Report 2012: Studienergebnisse zu Ursachen und Auswirkungen, aktuellen Praktiken, Herausforderungen und Strategien in Deutschland, Technische Universität München.

Liddy, E.D. (2001), "Natural language processing", In: *Encyclopedia of Library and Information Science*, 2nd Ed., Marcel Decker, Inc., NY.

Mehta, C. (2010), *Knowledge-based Methods for Evaluation of Engineering Changes*, Dissertation, University of Michigan, Michigan, 2010.

Morik, K., Jung, A., Weckwerth, J., Rötner, S., Hess, S., Buschjäger, S. and Pfahler, L. (2015), "Untersuchungen zur Analyse von deutschsprachigen Textdaten", Technische Universität Dortmund, Tech. Rep. 02/2015. https://doi.org/10.17877/DE290R-19161

Müller, R.M. and Lenz, H.-J. (2013), *Business Intelligence*, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-35560-8.

Mustafi, J. (2016), "Natural Language Processing and Machine Learning for Big Data", In: Mishra, B.S.P., Dehuri, S., Kim, E. and Wang, G.-N. (Eds.), *Techniques and Environments for Big Data Analysis: Parallel, Cloud, and Grid Computing*, Springer International Publishing, Cham, pp. 53-74. https://doi.org/10.1007/978-3-319-27520-8_4

Petersen, M. (2017). Considering sustainability in the development of consumer goods, DoctoralThesis Technische Universität Hamburg, 2017. https://doi.org/10.15480/882.1357

Potdar, P. and Jonnalagedda, V. (2018), "Design and development of a framework for effective engineering change management in manufacturing industries", *International Journal of Product Lifecycle Management*, Vol. 11 No. 4, p. 368. https://doi.org/10.1504/IJPLM.2018.097880

Rebentisch, E. et al. (2018), "Defining Agile Culture Using Topic Modeling", *Faculty of Mechanical Engineering and Naval Architecture*, The Design Society, Glasgow, UK, May 21-24, 2018, University of Zagreb, Croatia, pp. 2111-2120. https://doi.org/10.21278/idc.2018.0394

Riesener, M. et al. (2019), "*Literature Based Derivation of a Framework to Evaluate Engineering Change Requests*", PICMET 2019 Conference, Portland, Oregon, USA, August 25-29, 2019.

Röder, M., Both, A. and Hinneburg, A. (2015), "Exploring the Space of Topic Coherence Measures", In: Cheng, X., Li, H., Gabrilovich, E. and Tang, J. (Eds.), *WSDM'15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining,* January 31-February 6, 2015, Shanghai, China, Shanghai, China, 2/2/2015 - 6/2/2015 ACM Association for Computing Machinery, New York, NY, pp. 399-408. https://doi.org/10.1145/2684822.2685324

Roelleke, T. and Wang, J. (2008), "TF-IDF uncovered", *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, Singapore, Singapore, ACM, New York, NY, p. 435. https://doi.org/10.1145/1390334.1390409

Runkler, T.A. (2015), *Data Mining: Modelle und Algorithmen intelligenter Datenanalyse, Computational intelligence*, 2. Auflage, Springer Vieweg, Wiesbaden, https://doi.org/10.1007/978-3-8348-2171-3

Salton, G. and Buckley, C. (1988), "Term-weighting approaches in automatic text retrieval", *Information Processing & Management*, Vol. 24 No. 5, pp. 513-523. https://doi.org/10.1016/0306-4573(88)90021-0

Schuh, G. et al. (2019), "Data Mining Definitions and Applications for the Management of Production Complexity", *Procedia CIRP*, Vol. 81, pp. 874-879. https://doi.org/10.1016/j.procir.2019.03.217

Schuh, G. and Riesener, M. (2018), *Produktkomplexität managen: Strategien - Methoden - Tools, 3*, vollständig überarbeitete Auflage, Hanser, München. https://doi.org/10.3139/9783446443549

Schuh, G., Riesener, M. and Dölle, C. (2016a), "Concept for development project management by aid of predictive analytics", *2016 Portland International Conference on Management of Engineering and Technology (PICMET)*, Honolulu, HI, USA, pp. 2040-2047. https://doi.org/10.1109/PICMET.2016.7806640

Schuh, G. et al. (2016b), "Application of highly-iterative product development in automotive and manufacturing industry", *ISPIM Innovation Symposium,* The International Society for Professional Innovation Management (ISPIM), p. 1.

Sharafi, A. (2013), *Knowledge Discovery in Databases: Eine Analyse des Änderungsmanagements in der Produktentwicklung*, Zugl.: München, Techn. Univ., Diss., 2012, Springer Gabler Research, Springer Gabler, Wiesbaden. https://doi.org/10.1007/978-3-658-02002-6

Shearer, C. (2000), "The CRISP-DM Model. The New Blueprint for Data Mining", *Journal of Data Warehousing*, Vol. 5 No. 4.

Wänström, C. and Jonsson, P. (2006), "The impact of engineering changes on materials planning", *Journal of Manufacturing Technology Management*, Vol. 17 No. 5, pp. 561-584. https://doi.org/10.1108/17410380610668522

Wei, X. and Croft, W.B. (2006), "LDA-based document models for ad-hoc retrieval", *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, Seattle, Washington, USA, August 6-11, 2006, New York, NY, p. 178. https://doi.org/10.1145/1148170.1148204

Wickel, M. et al. (2015), "Comparison of Seven Company-Specific Engineering Change Processes", *Modelling and Management of Engineering Processes: Proceedings of the 3rd International Conference 2013*, Berlin, Heidelberg, s.l., Springer Berlin Heidelberg, pp. 125-136. https://doi.org/10.1007/978-3-662-44009-4_11

Wickel, M.C. (2017), *Änderungen besser managen. Eine datenbasierte Methodik zur Analyse technischer Änderungen*, Dissertation, Technische Universität München, München, 2017.

Wilberg, J. et al. (2015), "Using a Systemic Perspective to Support Engineering Change Management", *Procedia Computer Science*, Vol. 61, pp. 287-292. https://doi.org/10.1016/j.procs.2015.09.217