Health Informatics Meets eHealth
D. Hayn and G. Schreier (Eds.)

doi:10.3233/978-1-61499-759-7-1

Concept-Based Retrieval from Critical Incident Reports

Kerstin DENECKE[a,1]
[a]*Bern University of Applied Sciences, Bern , Switzerland*

Abstract. Background: Critical incident reporting systems (CIRS) are used as a means to collect anonymously entered information of incidents that occurred for example in a hospital. Analyzing this information helps to identify among others problems in the workflow, in the infrastructure or in processes. Objectives: The entire potential of these sources of experiential knowledge remains often unconsidered since retrieval of relevant reports and their analysis is difficult and time-consuming, and the reporting systems often do not provide support for these tasks. The objective of this work is to develop a method for retrieving reports from the CIRS related to a specific user query. Methods: atural language processing (NLP) and information retrieval (IR) methods are exploited for realizing the retrieval. We compare standard retrieval methods that rely upon frequency of words with an approach that includes a semantic mapping of natural language to concepts of a medical ontology. Results: By an evaluation, we demonstrate the feasibility of semantic document enrichment to improve recall in incident reporting retrieval. It is shown that a combination of standard keyword-based retrieval with semantic search results in highly satisfactory recall values. Conclusion: In future work, the evaluation should be repeated on a larger data set and real-time user evaluation need to be performed to assess user satisfactory with the system and results.

Keywords. Information Retrieval, Data Mining, Natural Language Processing, Critical Incidents Reporting.

1. Introduction

Access and retrieval of relevant information for patient safety and quality assessment is important in clinical contexts. The objective of critical incident reporting systems (CIRS) is to enable users, e.g. health care professionals working for a hospital, to report in an anonymous manner critical events that occurred in their working environment. Incident reporting has been instituted in healthcare systems in many countries for some time now, e.g. in Switzerland in 1997 [1], but not in all healthcare systems it is obligatory to report critical incidents. However, it has been shown that those anecdotal reports bear important information on limitations of systems and processes [2]. On the one hand, critical situations or even systematic errors can be identified by studying these reports which is crucial to develop countermeasures. On the other hand, once a measure to address certain problematic situations has been realized in a hospital, a database of incident reports could be used to check whether the measure was successful and the numbers of reported cases of a certain problem category dropped as expected. The reports allow to recognize trends relevant for further assessment.

In practice, reports are entered into CIRS by filling a digital reporting form. It contains multiple free text fields mainly asking for describing the problem or incident

1 Corresponding Author: Kerstin Denecke, Bern University of Applied Sciences, Quellgasse 21, Biel, E-Mail: kerstin.denecke@bfh.ch.

that has been recognized. Experiences from quality manager in hospitals showed that the free text fields bear the most important information for problem assessment. For identifying problems from these reports in a hospital, the entire database of free-textual event descriptions needs to be queried. So far, CIRS are not designed to support a retrieval of relevant reports matching a specific query or only structured fields can be queried. Thus, there is a substantial need for analyzing the free-textual parts of the reports automatically, identifying trends and frequently occurring incidents, very serious problems or even causes of critical incidents as the use case scenarios above show. The systematic access to big and complex data sets, especially in medical information systems is still challenging [3]. State of the art retrieval methods are based on methods like keyword matching, document indexing, scoring or term weighting [4], language models or inference networks [5]. In the last years, research in that area has focused on improving search results by taking domain knowledge into account [6] or using semantic methods [7]. In such techniques, words from both the query and the items in the corpus are mapped to concepts and relations in a knowledge source (typically an ontology), and retrieval is then based on semantic proximity in the background ontology.

Very limited work in natural language processing and information management considered medical incident reports. Akiyama et al. introduced a method to distinguish incident reports using artificial intelligence technology [8]. In more detail, characteristic words were extracted from incident reports, and co-occurrence networks of the characteristic words were created. Fujita et al. performed a linguistic analysis for incident reports in English [9]. They extracted characteristic words using natural language processing and they evaluated the degree of similarities between incident documents. In our work, we aim at handling the special requirements needed for processing incident reports from hospitals to support a search and navigation within a collection of reports. This use case differs significantly from many other search scenarios. As the size of the document collection under scrutiny is relatively small compared to a pool of abstracts of biomedical literature or an entire database of clinical documents comprising thousands of texts, a retrieval approach tuned towards a high recall is required. Objective of the retrieval is to get hints to potential problems in the workflow. The user values all facts that address a given information need and he or she would also accept a certain amount of false-positive results. The expected result are incident reports that match the query. The main contributions of the paper are: 1) Introduction and implementation of a concept-based retrieval method for incident reports in German, and 2) Evaluation and comparison of concept-based retrieval with a standard retrieval method.

2. Material and Methods

2.1. Requirements

In previous work, we studied user requirements and challenges of incident report retrieval [10]. In summary, a future incident report **retrieval system**: (1) provides a keyword search on the entire data set, (2) enables coordinated queries (AND, OR), (3) performs a semantic enrichment or automatic query expansion with synonyms (e.g. when only a drug name is mentioned in a text, the text should be anyway retrieved when searching for the keyword “medication”) and lexical variants, and (4) identifies matches that sound similarly (SoundEx, phonetic algorithm [17]).

Table 1: Examples of incident reports

Original incident message	Translation
Ein Pat. äussert Schmerzen. Kollegin bittet mich dem Pat. ein bestimmtes Schmerzmedikament zu spritzen welches bereits auf einem Tablett im Zimmer ist. Auf selbem Tablett befindet sich auch ein Nauseamedikament. Gebe dem Pat. dieses da es gleich aussieht und ich die andere Spritze nicht beachtet habe. Habe nicht nachkontrolliert was auf der Spritze steht.	One patient expresses pain. Colleague asks me to inject the patient a certain pain medication which is already on a tray in the room. On the same tablet there is also a nausea medication. Give the patient this because it looks the same and I did not pay attention to the other one. Have not checked afterwards what is written on the syringe.
Meine Kollegin bemerkt den Fehler kurze Zeit später.	My colleague notices the error a short time later.
Anästhesie hat das falsche Antibiotika ufgeschrieben bei der Postop verordnung (dieses Antibiotika hätte ein anderer Patient gehabt)	Anesthesia has put the wrong antibiotics at the postop prescription (this antibiotic would have had another patient)

2.2. Material

The basis of the analysis and retrieval experiment are 581 randomly selected incident reports from the Inselspital Bern. They originate from different clinics of the hospital and consist of at least of a date and a free-textual event description. Most of them have a title that is summarizing the critical event consisting mainly of one to three keywords. Sometimes, a potential measure for addressing the problem is suggested in a separate data field. An example is shown in Table 1.

2.3. Retrieval Methods

A corpus analysis [10] showed that the incident reports contain medical and non-medical named entities. We follow the hypothesis that this peculiarity requires a retrieval method that allows to search for medical concepts and for keywords. We apply Apache Lucene (https://lucene.apache.org/core/) to create word vectors for the documents in the data collection. Lucene is an open source, full-featured text search engine library written in Java. The standard settings are kept, meaning that the texts are tokenized, normalized to lower case, stop words are removed and stemmed. But, we are extending the search vector by semantic concepts. More specifically, query and reports are indexed by ID MACS® which results in a list of concepts of the Wingert nomenclature for each report and the query. The terminology server ID MACS® — *medical semantic network*, is a software provided by the German company ID Information und Dokumentation im Gesundheitswesen (http://www.id-berlin.de) [3,11].

When a document is analyzed, ID MACS® splits the text into its sentences. Each sentence is then broken up by a chunking method. The resulting chunks contain noun, verb or adverbial phrases. In each of these phrases the clinical and additional concepts are identified. If one potential word is found, it is mapped onto the respective concept of the Wingert Nomenclature. The latter is a German derivate of an early version of SNOMED [12]. It is a polyaxial nomenclature that contains ten axes of different categories of concepts. For example, the Topology-axis contains topological concepts, the Morphology-axis contains morphological concepts and the Procedure-axis contains concepts referring to medical procedures. In addition, the G-axis contains helpful concepts for certain adjectives and verbs and linguistic meta-information (e.g. „negated phrase“).

The surrounding words of an examined word are taken into consideration for concept mapping. Thereby, ambiguities can be resolved and inconvenient wordings can

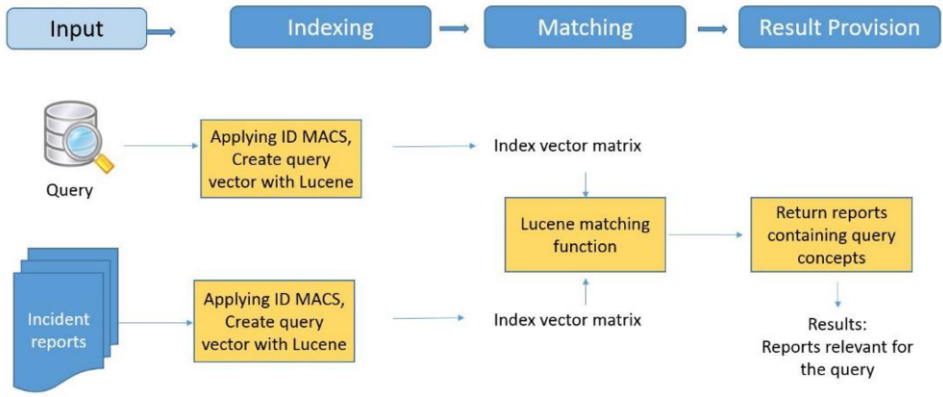

Figure 1. Processing pipeline of the semantic search

still be mapped correctly. In addition, ID MACS® handles synonyms or paraphrasing of certain entities or events and is able to map them correctly onto the corresponding concept. The algorithm's output is the text in a representative tree structure. The hierarchical structure of the tree represents the order of the individual phrases. The leaves contain the identified medical and linguistic concepts [13, 14].

This vector of indices is added to the Lucene generated index matrix. The retrieval returns documents that either contain the query term explicitly mentioned or a semantic concept referring to the query term. Figure 1 provides an overview on the retrieval method. We refer to this method by "combined". In the evaluation, we compare this combined method with a retrieval that only relies upon the concept match (referred to by "Semantic" in the following) and a retrieval that exploits the Lucene standard analyzer ("lucene").

2.4. Evaluation Strategy

In the evaluation, we are comparing these three retrieval methods. The objective is to determine the precision, recall and F-Score of the retrieval and to determine the differences and limitations of concept-based retrieval (combined and semantic) versus standard IR (lucene).

We are considering five topics for which documents are retrieved (see Table 2): delivery, drugs, prescription, control and signature. The topics and queries have been formulated by the future user. The queries that have been used for retrieval are listed in Table 2. The indicated search terms were concatenated by OR in our system. The gold standard has been created manually by a medical expert. He applied the query terms in

Table 2: Queries for the evaluation and the number of target documents according to the manual annotation

ID	Topic	Query terms	Translated query (not used in the evaluation)	No. of target documents
1	Delivery	Ausgabe, Abgabe, abgegeben	delivery, deliver	14
2	Drugs	Medikamente, Medikation	drugs, medication	284
3	Prescription	Verordnung	prescription, prescribe	176
4	Control	Kontrolle, kontrollieren	control	80
5	Signature	Visum	signature	1

the Excel search field and marked all relevant matches. We determined F-measure, precision and recall for the new introduced method (Combination of keyword-based and concept-based retrieval), but also for each approach on its own to be able to assess the differences in the retrieval methods. Errors or missing retrieval results have been assessed manually. Since for the retrieval task under consideration the recall is more important than the precision, we calculate also the F_2 – measure using (Equation 1)

$$F_\beta = (1 + \beta^2) * (precision * recall \,/\, \beta^2 * precision + recall) \quad (1)$$

with $\beta = 2$. This weights recall higher than precision.

3. Results

Table 3 shows the retrieval results for all queries and the three retrieval methods. It can be seen that the precision is lower for the semantic search and the combined approach than for the lucene retrieval method. Precision is in average 96% for the lucene retrieval, 51% for the semantic retrieval and 61% for the combined approach. The quality varies substantially for the semantic retrieval and combined approach depending on the query. The highest precision value of 94% is achieved with the semantic retrieval and combined approach for the query "Drugs". In contrast, the average recall of the combined approach is significantly higher with 96.5% than for the lucene and semantic approach. This means, many irrelevant documents are retrieved with the semantic and combined approach, but in particular for the combined method, the identified results mainly contain the relevant ones. The F2-Measure for the lucene retrieval and combined approach is similar with 0.78.

Table 3. Evaluation Results

Query ID	Query		Lucene	Semantic	Combined
1	Delivery	Precision	90%	44.8%	44.8%
		Recall	64.3%	93%	93%
		F2-Measure	0.682	0.765	0.765
2	Drugs	Precision	97.5%	94.7%	94%
		Recall	97.5%	94.4%	100%
		F2-Measure	0.975	0.944	0.987
3	Prescription	Precision	91.7%	65.1%	63.4%
		Recall	81.3%	96%	98.3%
		F2-Measure	0.831	0.873	0.885
4	Control	Precision	100%	52.8%	54.1%
		Recall	37.5%	82.5%	91.3%
		F2-Measure	0.4289	0.741	0.802
5	Signature	Precision	100%	0	50%
		Recall	100%	0	100%
		F2-Measure	1	0	0.5
	Average	**Precision**	96%	51%	61%
		Recall	76.1%	73.2%	**96.5%**
		F2-Measure	0.783	0.665	**0.789**

4. Discussion

In this work, a new retrieval method was introduced based on semantic term mapping. It has been shown that a combination of standard keyword-based retrieval with the semantic search results in highly satisfactory recall values. The lucene retrieval returns all texts containing the search term which can result in false positives since the context is not at all considered. The retrieval fails for the keyword-based approach lucene when the search term or a synonym is not explicitly mentioned in the text and semantic inferences would be necessary. For example, a report that only contains the term "heparin" would not be identified with a query like "medication". So far, the semantic relations included in the ID MACS® are not used, but could help in making such inferences. The semantic approach provides many false negatives due to the indexing process (i.e. the mapping of natural language to concepts of the ontology). Sometimes, the results are false positives, but for some cases these are true positives that were not determined using the Excel query. This holds true for query terms with synonyms. The semantic (and combined) search abstracts from lexical variants and synonyms and is thus more powerful than a simple keyword matching. For this reason, inflected verbs are mapped to the same concept (e.g. verordnet, Verordnung etc. (prescribe, prescription) are mapped to the concept referring to "Verordnung" (prescription)).Beyond the evaluation protocol, we recognized that proper names as query can fail in the semantic search. They cannot be indexed and will thus result in no retrieval results (e.g. ipdos). On the other hand, the keyword-based approach can fail for proper names given many writing variations of proper names (ipdos, i-pdos, i-p Dos, i-dos…).

Precision rates for information retrieval tasks in the biomedical domain have been assessed mainly from biomedical literature and achieved a precision between 70-90 % while the recall is around 70% [15]. Compared to this, the combined approach with semantic and keyword-based search results in a better recall (96%), and in a slightly lower precision. The idea of a semantic retrieval is not new, but still has not been considered for medical incident reports in German. The underlying terminology of ID MACS® with its semantic network provides a well suited resource for realizing the retrieval.

The approach was tuned towards a high recall. This results from the experience that the quality manager would rather go through some false positive documents instead of losing too much relevant information. From the current retrieval practice it is still a gain in time to have some irrelevant texts in the results set. In future work, the system should be tested with queries from multiple users. The queries used for the evaluation were to a certain extent artificial, since they contain only keywords as they are used in the existing retrieval method. Having a real retrieval system on hand could lead to more complex or syntactically incorrect queries.

The evaluation does not reflect the user satisfaction with the retrieval result. As next step, the retrieval system needs to be tested and used by quality manager and people accessing CIRS messages. It is obvious that time for retrieving relevant reports is decreasing using such retrieval methods. Therefore, we expect a high user satisfaction since they get support in analyzing the reports. With respect to the evaluation setting, it has to be mentioned that the dataset was relatively small and comprised only incident reports from one hospital. A ranking of search results was not considered, since it is irrelevant in the current search scenario.

To address the fact that specific queries cannot be easily formulated by the user, faceted search or data set visualization methods such as tag clouds could help. A first

assessment showed that tag clouds provide useful terms for further retrieval. We will consider this in our future works. Additionally, a query expansion using the semantic network of ID MCAS® will be considered. Query expansion is a representative technique of information retrieval. It generates alternative search terms or expanded queries on lexical or semantic level for improving the retrieval performance [16]. The retrieval can also be improved by including SoundEx technology [17] to retrieve also texts for query terms that sound similar to word in the reports.

5. Acknowledgement

We acknowledge Helmut Paula for providing and annotating the data. Further, thanks to ID Berlin for giving access to ID MACS® within the context of this work.

References

[1] Staender S, Daviers J, Helmreich B, Sexton B, Kaufmann M. The anaesthesia critical incident reporting system: an experience based database. Intern J Med Inform 47, 1997, 87-90

[2] Pham JC, Girard T and Pronovost PJ. What to do With Healthcare Incident Reporting Systems. Journal of Public Health Research, 2013, 2(3), 154-59

[3] Denecke K. Informationsextraktion aus medizinischen Texten. (Information extraction from medical documents). PhD Thesis. Shaker Verlag, Aachen, 2008.

[4] Manning CD, Raghavan P, Schütze H et al. Introduction to information retrieval, Volume 1. Cambridge university press, Cambridge, 2008

[5] Dakka W, & Ipeirotis PG . Automatic Extraction of Useful Facet Hierarchies from Text Databases. 2008 IEEE 24th International Conference on Data Engineering. doi:10.1109/icde.2008.4497455.

[6] Vit Novacek TG and Handschuh S. Coraal towards deep exploitation of textual resources in life sciences. Lecture Notes in Computer Science. Berlin/Heidelberg, 2009, 5651/2009:206–215

[7] Gonzalo J, Li H, Moschitti A, and Xu J. Sigir 2014 workshop on semantic matching in information retrieval. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '14, New York, NY, USA, 2014. ACM, 1296-1296.

[8] Akiyama M, Yamamoto S, Fujita K, Sakata I, Kajikaw Y. Effective Learning and Knowledge Discovery Using Processed Medical Incident Report. 2012 Proceedings of PICMET '12: Technology Management for Emerging Technologies, Vancouver, BC, 2012, 2337-2346

[9] Fujita K, Akiyama A, Park K, Yamaguchi EN, Furukawa H. Linguistic Analysis of Large-Scale Medical Incident Reports for Patient Safety. Studies in Health Technology and Informatics. Volume 180: Quality of Life through Quality of Information, 2012, 250 –254

[10] Denecke K: Automatic Analysis of Critical Incident Reports: Requirements and Use Cases. Stud Health Technol Inform. 2016;223:85-92.

[11] Kreuzthaler M, Bloice MD, Faulstich L, Simonic KM, Holzinger A. A Comparison of Different Retrieval Strategies Working on Medical Free Texts. JUCS, 2011, 17(7): 1109-1133.

[12] Wingert F. (Automated indexing based on SNOMED. Methods Inf Med, 1985, 24(1), 765-773.

[13] Denecke K, Bernauer J. Extracting Specific Medical Data Using Semantic Structures. AIM, 2007: 4594: 257-264

[14] Denecke K. Semantic Structuring of and Information Extraction from Medical Documents using the UMLS. Methods of Information in Medicine, 2008, 5(47), 425-34

[15] Ananiadou S, McNaught J. Text Mining for Biology and biomedicine. Artech House, Inc., Norwood, MA, USA, 2005

[16] Voorhees EM. Query expansion using lexical-semantic relations. In Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '94), W. Bruce Croft and C. J. van Rijsbergen (Eds.). Springer-Verlag New York, Inc., New York, NY, USA, 1994, pp. 61-69

[17] Zobel J and Philip Dart P. Phonetic string matching: lessons from information retrieval. In Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '96). ACM, New York, NY, USA, 1996, pp. 166-172