Finding Clusters of Similar Events within Clinical Incident Reports: A Novel Methodology Combining Case-Based Reasoning and Information Retrieval

Costas Tsatsoulis, Ph.D.[1] and Heather A. Amthauer, B.A.
Department of Electrical Engineering and Computer Science
The University of Kansas
Lawrence, KS 66045, USA
tsatsoul@ittc.ku.edu
tel: +1-785-864-7749
fax: +1-785-864-0387

Abstract

This paper discusses a novel methodological approach for identifying clusters of similar medical incidents by analyzing large databases of incident reports. The discovery of similar events allows the identification of patterns and trends, and makes possible the prediction of future events and the establishment of barriers and best practices. In our work we integrated two techniques from the fields of Information Science and Artificial Intelligence, namely Case-Based Reasoning and Information Retrieval, and achieved very good clustering accuracies on a test data set of transfusion medicine incident reports. Our work showed that clustering should integrate the features of an in-

[1] Corresponding author

cident captured in traditional form-based records, together with the detailed information found in the narrative included in event reports.

Keywords: clustering, clinical event reporting, data mining

Introduction

One of the goals of incident reporting systems is to allow their users to discover trends, identify patterns of organizational behavior, and predict future failures of the process. This is especially true for systems that collect reports across many organizations; such systems allow organizations to learn from the process shortcomings of others, and to correct their own operating procedures before similar errors appear locally.

To achieve these objectives the users of an incident reporting system should be able to point to a specific report and then query the system for other incidents that are *similar* to it. In essence, the users are interested in identifying a *cluster* of event reports that are exemplified by the report in question. The identification of the cluster provides valuable information to the users of an incident reporting system: how many reports are in the cluster, what is their distribution in time (which, in turn, helps in establishing trends), what are the exemplifying characteristics of the cluster, and so on.

Standard database retrieval cannot offer a measure of similarity; objects in a traditional database are accessed by exact matching of field values. While it is of some value to identify incident reports that have identical descriptions, it is a lot more probable that incident reports will be only *similar,* that is, will share common features, but will differ in other ones. Additionally, even features that are different in two reports may share some common characteristics; for example the incident time frame may be “4-8 am” in one report and “8-12 am” in another, but both times can be thought of as “morning.” When analyzing trends in medical incidents as well as when trying to identify best practices in response to incidents, medical personnel and quality assurance experts are interested in finding clusters of similar reports, that is reports that share some important common characteristics, instead of looking for identical reports. Similarity requires both a *syntactic* and a *semantic* matching of the features describing an incident report. Syntactic matching compares two strings of characters, for example “abc” and “Abc,” and determines if they are identical or, if not, how different they are (e.g. they differ in one out of three letters in the previous example).. Semantic matching compares two concepts, such as “male” and “man,” and determines if they represent the same thing or idea, and, if not, how close the two concepts are semantically. Case-Based Reasoning and Information Retrieval, two techniques from the field of Artificial Intelligence, offer tools to identify similar incident reports..

In this paper we first describe Case-Based Reasoning (CBR) and Information Retrieval (IR), and then describe our use of these techniques to identify clusters of similar documents in the MERS-TM (MEdical Reporting System - Transfusion Medicine) event reporting system, which is used to document incidents in transfusion services. When we applied CBR for the creation of clusters of similar reports we first identified the features of a transfusion incidence report that should be used as indexes (report descriptors useful in identifying similarity), we assigned different weights to each index as an indicator of its importance in establishing similarity, and defined domain-specific semantics to allow knowledge-based matching of indexes. In addition, we used techniques from information retrieval (IR) to analyze the textual description of the event that is attached to each report. We performed a set of experiments on a set of incident reports collected through the MERS-TM transfusion medicine incident reporting system [1], using CBR retrieval, IR retrieval, and also integrating IR with CBR. The goal of our experiments was to determine whether the CBR and IR retrieval methodologies alone would identify as similar cases that experts in transfusion services would also consider as such, and, whether a combination of CBR and IR retrieval would have superior retrieval performance than either technique alone. The results of each retrieval, clustering, and similarity assessment were evaluated with the help of experts in the area of quality assurance in transfusion medicine who calculated the number of false positives and negatives in the clusters of similar incident reports generated by our software.

Our results indicate that the integration of CBR with IR improves performance of the retrieval system and offers good recall and accuracy.

Application of Case-Based Reasoning and Information Retrieval to Incident Reports from Transfusion Services

The MERS-TM incident reports were analyzed by experts in the field of transfusion services who defined a subset of the report features that should be used as indexes in our CBR system. Some of these features are the discovery time, the discoverer's job description, the point in the process the event was discovered, where it first occurred, the causal and antecedent codes, and so on. The experts also assigned a weight between 1 and 5 to each index, where the higher weight indicates greater importance of a feature in matching and clustering. For example, where an event first occurred was weighted as a 5, while the time an event was discovered was given a 1, and the discoverer's job description a 3.

For some attributes the experts gave us conditional weights. For example, a causal code would receive a weight of 1 or 2, depending on whether it was based on a rough examination of the incident or on an in depth analysis.

The experts also defined hierarchies of attribute values that allowed us to define partial matches. For example:

1. For the attribute indicating when an incident was discovered, the values in the pairs (12-4 am, 4-8 am), (8-12 noon, 12-4 pm) and (4-8 pm, 8-12 midnight) were considered partially similar. So, for example, a report with value "8-12 noon" would have a partial match with a report with value "12-4 pm."
2. For the attribute indicating the job description of the person discovering the incident, the values in the following sets were considered as matching partially: (RN, LVN/LPN), (Staff, MLT, MT, QA/QC, RN, LVN/LPN), (Supervisor, MT, QA/QC,RN), (Resident, MD/DO), and (MLT, MT, QA/QC).

and so on.

In our CBR system we only had a single level of hierarchy of feature values, and every partial match was assigned a value of 0.7 (a perfect match received a value of 1.0 and a non-match a value of zero).

Our approach to the IR portion of this study uses the vector-space model (VSM) and the cosine comparison measure, as described above. In our case, a document is considered the free-text of the report portion that describes what happened. The removal of noise from the text was difficult due to the domain specific abbreviations used. For example, “OR” was used mostly as an abbreviation for “operating room,” not as a conjunction. So as not to lose important abbreviations, no stop words were removed. Matching based on words that do not carry a lot of meaning due to their high frequencies is easy to identify, so the non-removal of stop words is easily handled.

Next, we performed a set of experiments to establish the efficacy of CBR and IR in clustering similar clinical incident reports. For the experiments we used a MERS-TM dataset of approximately 600 reports collected by the transfusion services of two hospitals and graciously made available to us by the MERS-TM group led by Dr. Harold Kaplan of the Presbyterian Hospital of Columbia University in New York, USA. The incident reports were indexed for CBR retrieval as indicated above, and also pre-processed for IR retrieval. After the incident reports were indexed they were entered in a “case base,” that is, in a storage file that makes comparisons and similarity assessment possible through our software. Similarly, after IR pre-processing, the incident reports were stored in a structure appropriate for IR retrieval.

The goal of our experiments was to determine: 1. whether CBR retrieval would identify as similar cases that experts in transfusion services would also consider as such; 2. whether IR retrieval would do the same; and, 3. whether a combination of CBR and IR retrieval would have superior retrieval performance than either technique alone. As a baseline test, we performed retrieval using equal weights for all indexes; the goal was to establish whether the index weights given to us by the experts improved CBR retrieval, or not.

We randomly selected 24 cases out of the approximately 600 incident reports in the dataset (to avoid confusion with the cases in the case base, the cases we used to match against will be called "reports" from now on) and for each of these reports we retrieved the 10 most similar cases from the case base we created from the processed incident reports. This experiment attempted to establish the usefulness of CBR for finding clusters of similar medical incident reports (goal 1 of our experiments, as described above). An example of two matching transfusion incident reports is shown in figure 3. In the figure, we show parts of two reports with color-coded matching and non-matching values, and an overall matching value of 0.49.

We used the same 24 reports and identified similar ones using only an IR-based keyword match of the text included with each case. This experiment attempted to establish the usefulness of IR for finding clusters of similar medical incident reports (goal 2 of our experiments, as described above).

Finally, we combined the results of the CBR and IR retrieval as follows: We assigned to the matching percentage of each retrieval technique a weight between 0.9 and 0.1, in increments of 0.1, making sure that the sum of the two weights always equaled 1.0. In other words, the CBR match value was weighted by 0.9, 0.8, 0.7, …, 0.2, 0.1, while the IR match value was weighted by 0.1, 0.2, 0.3, …, 0.8, 0.9. The goal of these combined experiments was establish the whether the combination of CBR and IR retrieval would offer superior performance for finding clusters of similar medical incident reports (goal 3 of our experiments, as described above).

The cases were then re-ranked based on the new combined weight, resulting in nine new rankings. In our base line test we performed CBR retrieval using equal weights for all indexes (the weights were set to 1, since the similarity value is normalized). The result of all these experiments was 12 sets of ranked cases, which were similar to the original report (CBR only, IR only, CBR with no weights, and nine rankings with varying weights assigned to the CBR and IR similarity values). In figure 4 we show a flow chart of operations, starting with the pre-processing of the MERS-TM incident reports and ending with the expert evaluation of the CBR and IR clustering.

For evaluation, for each report, we collected the top five of the retrieved cases from each experiment. Since many of the retrieved cases for the different experiments were the same, the result was a set of between 10 and 20 cases for each report. To these cases we added one randomly selected case from the database, to use as a control point for the evaluation. These cases were ordered randomly, so as not to give any hint to the evaluators. The two experts who participated in the evaluation of our work were Ms. Barbara Rabin Fastman of Columbia University's New York Presbyterian Hospital and Ms. Quay Mercer, currently with the University of Texas Southwestern Medical Center at Dallas. Both are experts in transfusion services and quality assurance of medical and hospital processes. Our experts were asked to evaluate whether the cases matched the report or not on a four-point scale: "Almost Identical," "Similar," "Not Very Similar," and "Not Similar At All." This scale is clearly subjective, and its intent is to give the experts the freedom to express their personal opinion of the quality of the performance of the similarity algorithm without having to understand how the algorithm works.

The experiments and the evaluation of their results were performed during 2002 and early 2003. The incident reports were handled in an electronic format (transformed appropriately for CBR and IR as described above), and the CBR and IR clusterings were performed using software we developed. The evaluation of the results by the experts was analyzed by statistical software to summarize it and to allow us to draw generalized conclusions.

Analysis of Experimental Results

We analyzed the results of the system and the experts' evaluation in the following manner:

1. We classified all cases ranked by the experts as "Almost Identical" and "Similar" as "Retrievable," while the other two rankings indicate cases that should be "Non-retrievable."
2. We studied the results of the 12 experiments (CBR only, IR only, CBR with no weights, and nine rankings with varying weights assigned to the CBR and IR similarity values) for different similarity matching thresholds (ranging from 0.1 to 1.0 in 0.1 increments). These thresh-

olds indicate what cases should be added to the cluster of similar ones. For example, a 0.4 threshold would include in the cluster cases which match with similarity value of 0.4 and above.

3. Our two quality criteria were *recall* and *accuracy*, which is what percentage of the retrievable cases we did retrieve, and what percentage of the non-retrievable cases did we *not* retrieve[2]. In other words, recall tells us how many of the appropriate reports we are finding, while accuracy tells us how many of the inappropriate reports we are avoiding (one minus accuracy would give us the percentage of the incorrect reports we are including in our similar cluster, indicating false positives). Clearly, we want high recall and accuracy.

We expected that as the similarity threshold was raised, recall would be lower and accuracy would improve: a lower similarity threshold would assume that most cases were similar and as a result include all the retrievable ones, but also many non-retrievable; as the threshold is increased, fewer cases are considered similar, excluding some retrievable ones, but, hopefully, also excluding most non-retrievable ones, too. We also expected that our CBR system would do better than the CBR with no weights, since the weights were assigned by experts specifically to assist in matching and similarity assessment.

We had no expectations about the performance of the integrated CBR and IR retrieval, since no similar experiments had been performed in the past.

Some of the results of our experiments are shown in Tables 1 to 5. As expected, as the matching threshold is increased, recall is lowered but accuracy increases greatly (Table 1). The asterisks indicate that no cases were retrieved that were above the listed matching thresholds. Clearly, the CBR-only retrieval does very well with recall, but poorly with accuracy. This may be an indication that the report fields used as indexes are superficial descriptors of an event, and as such do not offer the detail necessary to distinguish between dissimilar reports.

We next compared the recall and accuracy of the CBR system using expert-assigned weights versus the CBR system using equal weights. In Table 2 we are listing the difference in the quality of recall and accuracy as a function of the matching threshold. The asterisks indicate that no cases were retrieved that were above the listed matching thresholds. As expected, the recall of the CBR system with weights is substantially better than the one of the CBR system with equal weights. Table 2 would seem to indicate that CBR with equal weights has a better accuracy, but closer inspection of the results showed this

[2] These are also known as “true positives” and “true negatives” respectively.

not to be the case, since CBR with equal weights classified almost *all* cases as not similar, and, thus, would trivially exclude non-retrievable ones.

We also examined the accuracy and recall of the IR retrieval as a function of the matching threshold. As expected, as the matching threshold is increased, recall is lowered but accuracy increases greatly. The problem with IR retrieval is that the drop-off in recall is extremely steep. Our hypothesis is that the text in the MERS-TM reports stresses case-specific details that allow differentiation between dissimilar ones, but also precludes the identification of similar ones that may share more general characteristics (Table 3).

We next performed similarity retrieval using a weighted combination of IR and CBR, and the results are shown in Table 4 (Note that we are only listing results where recall and accuracy were above 70%). Interestingly, the best results occur in lower matching thresholds, and when –in general- the IR retrieval's contribution is greater or even dominant.

In general, the CBR system had better recall but worse accuracy than the IR system. The integration of CBR with IR produced the best results, since it combined the strengths of both techniques. Figure 5 shows the sum of accuracy and recall plotted against the matching threshold for 11 experiments (the

CBR with equal weights is not included since it was used only as a baseline test). The best combined recall and accuracy values happen for matching thresholds of 0.40 and 0.50, and for combined, weighted CBR and IR retrieval.

Since CBR seemed to identify retrievable cases well, and IR seemed to identify non-retrievable cases with over 90% accuracy, we performed one more experiment to examine an integrated CBR and IR system where each technique is used independently and then their results combined, to exploit each method's strength. Specifically, we conducted six additional experiments as follows: we performed CBR retrieval using the thresholds where CBR gave its best recall result, namely at 0.7, 0.6 and 0.5. Next we performed IR retrieval using 0.5 and 0.4 as thresholds, where IR had the best accuracy. We then created the six possible intersections of the three CBR and two IR sets. Table 5 summarizes these results. As can be seen these results are not substantially better over the ones achieved in the previous experiments, although they have outstanding recall and, in one case (CBR 0.6; IR 0.5), very good accuracy.

Discussion

The goal of our work was to examine, first, whether CBR and IR are useful in finding clusters of similar clinical incident reports, and, second, whether the combination of the two techniques offers improved clustering. For this we used a database of approximately 600 incident reports of transfusion services collected through the MERS-TM system. The clustering was done by software we developed, and the evaluation was performed by two experts in quality assurance of transfusion and other clinical services.

Our experiments showed that CBR is useful in identifying similar medical incident reports, but has accuracy problems. It seems that a lot of the detail of a case is contained in the textual description provided by the reporters of the event, and this is demonstrated by the retrieval accuracy of IR. On the other hand, the text in the reports is too detailed to provide sufficient abstract descriptions, leading to good accuracy but poor recall for IR-only retrieval. The combination of CBR and IR techniques, either as a weighted sum of similarity values or as the intersection of separate trials, greatly improved accuracy and recall. Based on our results we recommend very strongly that future systems that are developed to cluster medical event reports integrate both the field values and the text of the reports in their methodological approach.

There is very little other research that has examined the combination of CBR and IR for clustering, and it tends to support our findings. Specifically, the DRAMA system used text to enhance its case-based reasoning process by analyzing free-form text that was part of aircraft design documents to capture rationale and interrelationships of design choices. In an example presented in the paper the information in the text associated with a case improved retrieval, but the authors did not provide a systematic evaluation of the integration of the technologies [6].

There are two directions that our future work could take: first, the analysis of a large corpus of incident reports, and, second, the theoretical and experimental analysis of the best combination of CBR and IR. Our sample of incident reports (approximately 600 in total) is small compared to the size of databases of medical reports that are being created globally. It would be interesting to study how the size of the underlying database of reports affects the performance of clustering. Also, our work has provided some indications that CBR and IR work best when combined; future work should examine under what circumstances each technique offers the best benefit (e.g. more detailed versus more abbreviated text), and how the two techniques can be best combined to provide optimal clustering and retrieval results.

Acknowledgements

We would like to thank Barbara Rabin Fastman from Columbia University's New York Presbyterian Hospital, and Quay Mercer currently with University of Texas Southwestern Medical Center at Dallas, for assisting in the evaluation of our system. We would also like to thank the anonymous reviewers for improving the exposition of our work. The work describe in this paper was supported in part by grant RO1 HL53772 from the National Heart, Lung and Blood Institute of the National Institutes of Health, through a subcontract from Columbia University.

References

1. Kaplan HS, Callum JL, Rabin Fastman B, et al. The Medical Event Reporting System for Transfusion Medicine (MERS-TM): Will it help us get the right blood to the right patient? Transfusion Medicine Reviews. 2000; 16(2): 86-102.
2. Kolodner J. Case-Based Reasoning. San Mateo, CA: Morgan Kaufmann. 1993.
3. Salton G. The SMART Retrieval System – Experiments in Automatic Document Processing. Englewood Cliffs, NJ: Prentice Hall, 1971.

4. Salton G, McGill MJ. Introduction to Modern Information Retrieval. New York : McGraw-Hill, 1983.
5. van der Schaaf TW. Near Miss Reporting in the Chemical Process Industry. Eindhoven, Netherlands: Technical University of Eindhoven, 1992.
6. Wilson DC, Bradshaw S. CBR Textuality. Expert Update 2000; 3(1): 28-37.

Case-Based Reasoning (CBR)
Case-Based Reasoning is a problem solving paradigm that is based on psychological theories of human cognition and provides the foundations for a technology for intelligent systems [2]. Avoiding the details of the theory Case-Based Reasoning, we can describe it as based on the intuitive notion that human expertise is not based on rules or other formalized structures, but on experiences. Human experts differ from novices in their ability to relate problems to previous ones, to reason based on analogies between current and old problems, and to use solutions from old experiences. The process of reasoning using experiences or cases can be described by the following steps: **1. Retrieve:** Given a new problem, retrieve a similar past case from memory. The past case contains the prior solution. **2. Modify:** The old solution is modified to conform to the new situation, resulting in a proposed solution. **3. Test:** The proposed solution is tested for successful solution of the current problem. **4. Learn:** If the solution fails, explain the failure and learn it to avoid repeating it. If possible, repair the failure, generate a new proposed solution and re-

turn to step **3.** If the solution succeeds, incorporate it into the case memory as a successful solution, and stop.

Since our work concentrates on retrieval, our discussion here will be constrained to this part of a CBR system.

A CBR system must select the best case or cases from memory. The question that must be answered is what constitutes an *appropriate* or *similar* case. What are the criteria of closeness or similarity between cases, and how should cases be indexed? Indexing a case is essential in establishing similarity, since the indexes help define the elements of a problem that are important.

During retrieval each case must be compared to the current problem, and be assigned a degree of similarity. Then the retrieving program will select the cases with the highest degree of similarity. Consequently we need to define what we mean by "best match" or, as usually called in conceptual retrieval, what we mean by "similar(ity)". The simplest method would be to look at *structural* or *syntactic similarities* between the current problem and a case. This demands an exact match between index values, in a manner identical to a database retrieval. (Note that this is a simplification of structural matching. One can demand a perfect syntactic match only of symbolic values (i.e., nonnumeric ones); the same is not true for numerical ones. For numbers a perfect match may be based on a formula; e.g. " x is qualitatively equal to y if it is y±20%.") If two values match structurally we say that they match *perfectly*(or, if we wanted to assign a degree of match between 0 and 1, where 0 is absolute mismatch, a structural match would receive a value of 1.0). For example, we would day that “ABC” and “ABC” match perfectly (they are structurally identical, that is, they look the same), while “ABC” and “DEF” do not match, since they don’t look the same at all. On the other hand, we could define partial similarity, and say, for example, that “ABC” matches “XBC” with 67% match, since the two strings share two out of three letters.

Deciding whether two values match or not can also lead to a partial (or *semantic*) match. The concepts represented by the case indexes are placed on a hierarchy of classes and their subclasses. For example, one may say that "beef" and "chicken" are subclasses of "meat." Then, "beef" and "chicken" match partially since they are different concepts, but they are both subclasses of the superclass "meat." We can assign a value to this partial match based on the level of the hierarchy where values match. For example, a complete match can be given 1.0, and for moving up a level of the hierarchy we may want to multiply the match by 0.7 (1, 0.7, 0.5, 0.35, ...). Creating a membership hierarchy is just one way to establish partial similarity of symbolic values. Some of the similarity can be rule-based, where the rules are defined by experts. For example, an expert can give a rule that says that "the emotional state of anger is similar with degree 0.8 to the emotional state of rage," etc.

Indexes can be assigned a weight (in an arbitrarily selected scale) that indicates the contribution of a particular index to establishing similarity. Usually, index weights are assigned by domain experts who are best suited in estimating which characteristics of a case are the most relevant ones.

After we determine which index values are qualitatively similar or equal, we compute a similarity value for the whole case. Usually this is done in a nearest neighbor method, which is a weighted average. For example, we can compute the degree of similarity as:

$$similarity = \frac{\sum w_i \times sim(f_c, f_p)}{\sum w_i}$$

where w_i is the weight for a matching feature, and *sim* is the degree of match between the old case f_c and the current problem f_p.

Information Retrieval (IR)

Information Retrieval (IR) systems are used for indexing, searching and recalling text or other unstructured forms of data. IR's primary basis for text retrieval is through the use of weighted keywords. Since IR systems do not require any domain-specific knowledge, IR systems can be applied in any domain where textual documents are available.

Traditionally, text documents are pre-processed, where common words (or “stop words”), such as “a,” “and,” “the,” etc. are removed from the document. Next *stemming* is performed, where words are reduced to their stem, so that, for example, “independence” and “independent” are represented by the common stem “independ.” Next, the text tokens are stored in a structure that allows quick comparison and retrieval.

One approach in IR for document retrieval is the vector-space model (VSM) [3], where each document is represented by a list (vector) of terms. These terms have associated weights that describe a term’s value for a document. The weighting system for each term in the document uses a tf-idf scheme. (tf = term frequency; idf = inverse document frequency). In this term-weighting scheme, the tf and idf are calculated in the following manner:

tf = frequency of the term in the document/ the frequency of the most frequent word in the document

idf = $\log_{10}$(total number of documents in the collection/ the number of documents in the collection that contain the term)

Thus the weight of a term is calculated by:

Weight = tf * idf

Using the VSM makes it possible to compare two documents using vector algebra, as, for example, the cosine measure of similarity [4]. With this method, the degree of similarity between two documents is determined by the cosine of the angle between the vectors that represent the two documents (the smaller the angle, the more similar), so that a document might be retrieved even if it shares only a few terms.

A Description of the MERS-TM Incident Reporting System for Transfusion Medicine

The Medical Event Reporting System for Transfusion Medicine (MERS-TM) is an event reporting system developed for transfusion services and blood centers to collect, classify, and analyze events that could potentially compromise transfusion safety [1]. The incident reports of MERS-TM consist possibly of three parts. The first two parts are mandatory; one of them describes the incident with a set of surface features, such as the time and date the incident was discovered, by who it was discovered, when it occurred, location code of the point of occurrence, and so on. The other mandatory document is the QA investigation report, which includes codes describing the causal events (MERS-TM uses the Eindhoven classification system for causes of events [5]), any preventive actions taken, and the type of investigation conducted. In addition to the surface features and the causal event codes, the MERS-TM incident reports always include a brief (1-2 lines) textual description of the event. If the organization decides to perform a detailed investigation, it will generate the third, optional part of the report, which includes detailed information about the consequent and antecedent events.

Figure 1 shows the information that a user may enter to describe in MERS-TM the discovery of a transfusion medicine incident. Figure 2 shows part of a completed detailed investigation report entered in MERS-TM, displaying the causal codes based on the Eindhoven classification system [5].

Key Messages
• Databases of medical incident reports need to become *active* and provide answers instead of simply history. One way to do so is to identify clusters of similar incident reports that help in determining patterns, trends, and best practices.
• Case-Based Reasoning offers a useful methodology for identifying similar medical incident reports, but has accuracy problems
• The integration of Case-Based Reasoning and Information Retrieval greatly improves the recall and accuracy in clusters of similar medical incident report

Future Research and Development
• Analyze a larger (greater than 5,000) corpus of incident reports to determine the effects of the data size to the results of similarity clustering
• Study the optimal combination of Case-Based Reasoning and Information Retrieval in the creation of clusters of similar incident reports.

TABLES

Table 1: Analysis of the clustering quality of CBR by examining recall and accuracy over different matching thresholds

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Recall	1.00	1.00	1.00	1.00	1.00	0.95	0.76	*	*	*
Accuracy	0.00	0.00	0.00	0.00	0.00	0.05	0.57	*	*	*

Table 2: Comparison of CBR vs. CBR clustering quality as the difference of recall and accuracy of the two techniques. Positive numbers indicate better performance.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$Recall_{CBR}$-$Recall_{CBRequal}$	0.05	0.15	0.21	0.30	0.45	0.57	0.50	*	*	*
$Accuracy_{CBR}$-$Accuracy_{CBRequal}$	-0.35	-0.53	-0.73	-0.81	-0.88	-0.91	-0.40	*	*	*

Table 3: Analysis of the clustering quality of IR experiments examining recall and accuracy over different matching thresholds

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Recall	1.00	1.00	1.00	0.96	0.79	0.52	0.38	0.25	0.22	0.22
Accuracy	0.00	0.00	0.11	0.46	0.64	0.89	0.93	0.96	1.00	1.00

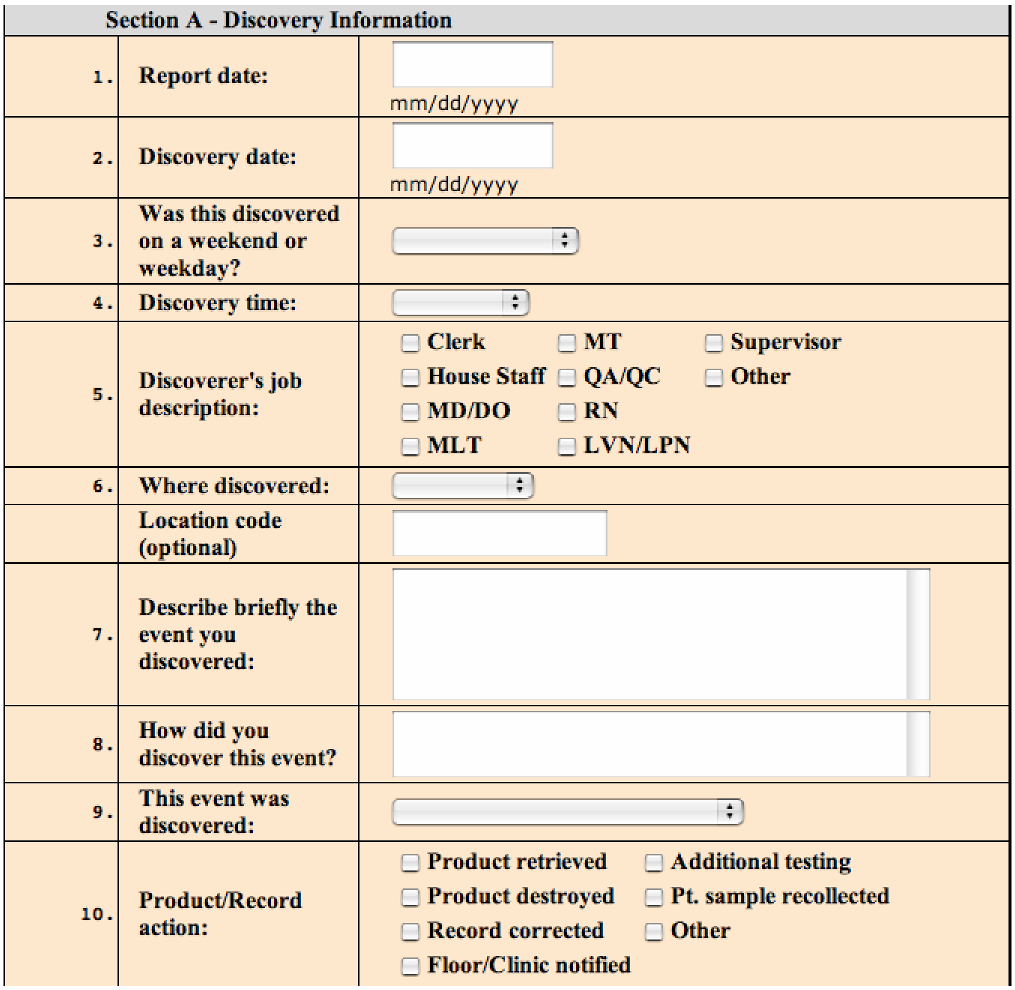
Table 4: Best recall and accuracy results achieved by combining CBR with IR

	CBR+IR (10:90) Thresh.=0.3	CBR+IR (40:60) Thresh.=0.4	CBR+IR (30:70) Thresh.=0.4	CBR+IR (20:80) Thresh.=0.4	CBR+IR (10:90) Thresh.=0.4	CBR+IR (60:40) Thresh.=0.5
Recall	0.79	0.80	0.75	0.73	0.79	0.75
Accuracy	0.73	0.75	0.80	0.78	0.73	0.80

Table 5: Results of integrating CBR with IR retrieval for the best retrieval thresholds of each technique

	CBR 0.7 IR 0.5	CBR 0.7 IR 0.4	CBR 0.6 IR 0.5	CBR 0.6 IR 0.4	CBR 0.5 IR 0.5	CBR 0.5 IR 0.4
Recall	1.00	0.99	0.92	0.93	0.93	0.88
Accuracy	0.13	0.17	0.66	0.25	0.40	0.52

FIGURES



Section A - Discovery Information		
1.	Report date:	mm/dd/yyyy
2.	Discovery date:	mm/dd/yyyy
3.	Was this discovered on a weekend or weekday?	
4.	Discovery time:	
5.	Discoverer's job description:	Clerk, MT, Supervisor, House Staff, QA/QC, Other, MD/DO, RN, MLT, LVN/LPN
6.	Where discovered:	
	Location code (optional)	
7.	Describe briefly the event you discovered:	
8.	How did you discover this event?	
9.	This event was discovered:	
10.	Product/Record action:	Product retrieved, Additional testing, Product destroyed, Pt. sample recollected, Record corrected, Other, Floor/Clinic notified

Figure 1: The "Discovery Information" section of MERS-TM. Here the user records how the transfusion medicine incident was discovered.

	Report accession number	100
1.	**Consequent (discovery) code:**	1 AV
2.	**Antecedent (1st occurrence) code:**	US
3.	**Significant antecedent (occurrence) code:**	OE
4.	**Additional description of event (optional)**	
5.	**Risk Assessment:**	**QES** .10 **QEP** .50 **Final RAI 0.25**
6.	**Organizational risk?**	None
7.	**Follow up:**	☑ **Propose action** ☐ **Consider action** ☐ **Monitor** ☐ **External report to other dept/org** ☐ **FDA Reportable**
8.	**If appropriate, describe the long-term preventive action to be taken:**	

Figure 2: Part of a completed detailed investigation report from MERS-TM. The quality assurance persons performing the investigation have identified and recorded a variety of causal and risk codes.

Rank:1 - Case ID: 1999100078 - Similarity: 0.491525

Expanded Similarity: 0.210526

Attributes in Retrieved Case and in CurrentCase

Attribute	Entered Case Value	Retrieved Case Value
Report_Accession_No	1999100080	1999100078
Report_Date	1999-03-11 00:00:00	1999-03-09 00:00:00
Discovery_Date	1999-03-11 00:00:00	1999-03-09 00:00:00
Discovery_Time	4-8 PM	4-8 PM
Discovered_By	Trans. Service	Trans. Service
Discoverers_Job_Description	MLT	MLT
Discovers_Dept	Trans. Serv.	Trans. Serv.
What_Happened1	DIRECTED PLASMA PUT IN	SUNNYBROOK AND CBS NUMBER
What2	REGULAR INVENTORY	ASSIGNED TO UNIT
What3		
How_Discovered1	CBS CALLED TO WARN OF	AT TIME OF ABO TYPING OF
How2	ARRIVAL AFTER RECEIPT	DONOR UNIT
How3		
Point_in_Process_Discovered	Other	Other
Date_Event_Occurred	1999-03-11 00:00:00	1999-03-09 00:00:00
Where_Event_Occurred	Other	Other
Occurrence_Time	4-8 AM	8-12 Noon
Job_of_Person_Involved	MLT	MLT
Product_Record_Action	Unit retrieved	Pt. record corrected
Investigation_Type	Expanded Investigation	Expanded Investigation

Figure 3: Example of matching of two incident reports. Attributes in black are not used in matching; attributes in green match perfectly; attributes in red do not match. The overall matching value of these two reports in 0.49.

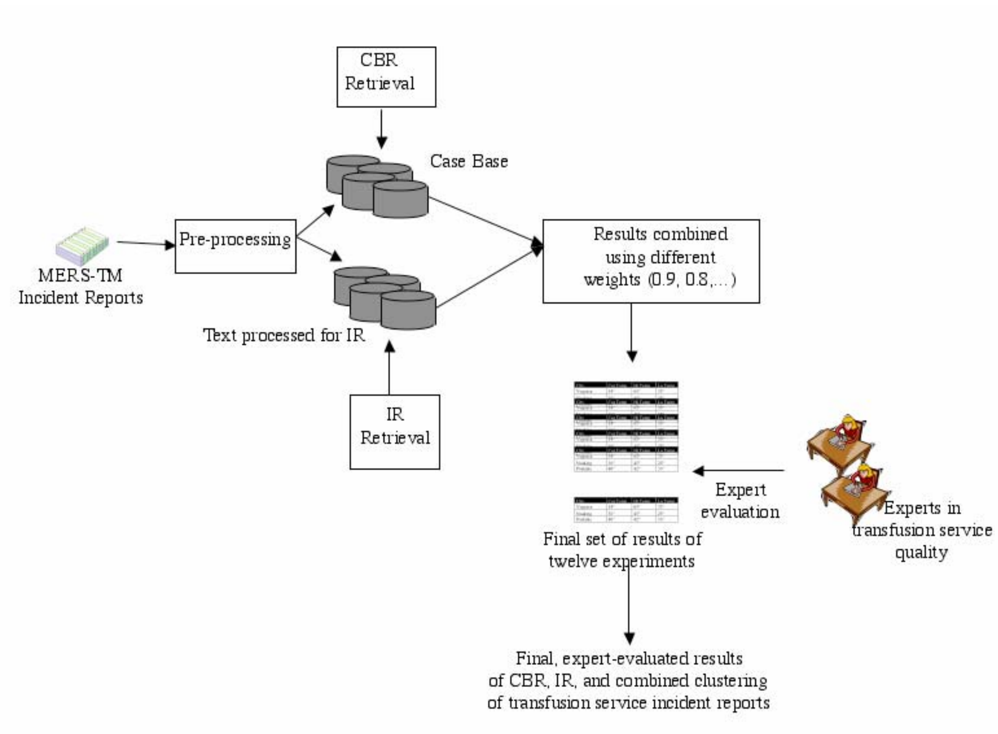


Figure 4: Flow chart of operations of the system.

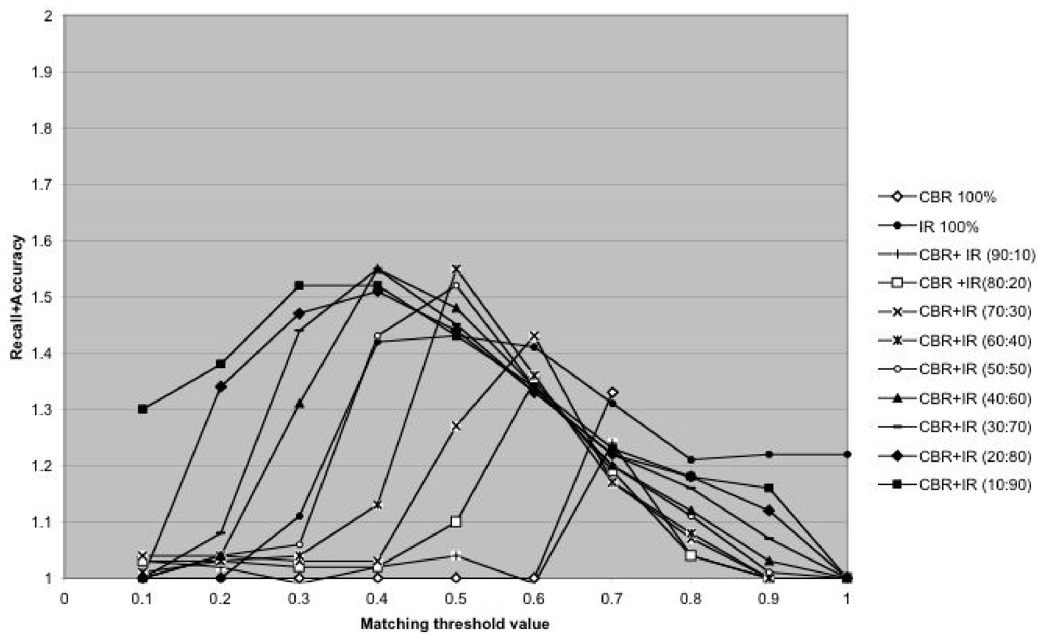


Figure 5: Plot of the sum of recall plus accuracy versus the matching threshold for CBR, IR and combined weighted CBR+IR experiments ran.