Summarisation and visualisation of e-Health data repositories

Conference or Workshop Item

Summarisation and Visualisation of e-Health Data Repositories

Catalina Hallett, Richard Power, Donia Scott
Computing Research Centre
The Open University
{C.Hallett, R.Power, D.Scott}@open.ac.uk

Abstract

At the centre of the Clinical e-Science Framework (CLEF) project is a repository of well organised, detailed clinical histories, encoded as data that will be available for use in clinical care and in-silico medical experiments. We describe a system that we have developed as part of the CLEF project, to perform the task of generating a diverse range of textual and graphical summaries of a patient's clinical history from a data-encoded model, a *chronicle*, representing the record of the patient's medical history. Although the focus of our current work is on cancer patients, the approach we describe is generalisable to a wide range of medical areas.

1 Introduction

Records of cancer patients are very rich: in addition to a thousand or more numeric data points arising from successive laboratory tests and a chronology of five or six hundred significant events – such as the dates tests were requested or performed, clinics attended or drugs dispensed – our typical patient files will also contain between fifty and a hundred and fifty narrative clinic letters, together with a similar number of reports interpreting a variety of investigations (e.g., Xray, body scan, etc.).

The computer readable part of an electronic patient record for direct clinical care is, therefore, a record of multiple events with no explicit semantic links between them: it records most of what was done, but very little of why. As a result, much if not most of the valuable clinical information remains machine unreadable, locked within the narrative letters and reports exchanged between doctors.

One of the aims of the Clinical e-Science Framework (CLEF) project (Rector et al., 2003), under which the research reported here is being conducted, is to establish a technical infrastructure for managing research repositories of aggregated patient data arising from routine medical care across potential multiple sites and institutions, in support of biomedical research. Information is extracted from medical narratives[1] and aggregated with structured data in order to build complex images of a patient's medical history which model the story of how patient illnesses and treatments unfolded through time: *what* happened, *when*, *what* was done, *when* it was done and *why*. The resulting complex semantic network, termed by us a *chronicle*, allows the construction of targeted summarized reports which do more than present individual events in a medical history: they present, in coherent text, events that are semantically and temporally linked to each other.

This paper discusses the problem of presenting aggregated clinical data: assuming the full richness of clinical information could be made available – whether extracted from clinical records in their current form or acquired a priori using an entirely different data capture paradigm (e.g., structured data entry) – how might that information be represented and exploited for the maximal benefit of clinical research and clinical care? Of particular interest to us here is the problem of automatically generating targetted and comprehensible textual reports from the data-encoded view of a patient's medical history.

In presenting medical histories we are trying to circumvent the shortcomings of textual reports

[1] Using Natural Language Processing techniques, see (Harkema et al., 2005).

by combining them with visual navigation tools. In this way, we take advantage of the better accesibility and interactivity offered by visual timelines as well as of the ability of natural language to convey complex temporal information and to aggregate numerical data.

2 Types of report

The intended end-user of the generated reports is a GP or clinician who uses electronic patient records at the point of care to familiarise themselves with a patient's medical history and current situation. A number of specific requirements arise from this particular setting:

- Events that deviate from the norm are more important than normal events (e.g., an examination of the lymphnodes that reveals lymphadenopathy is more important than an examination that doesn't). However, normal events should also be available on demand.
- Some events are more important than others and they should not only be included in the summary but also highlighted (through linguistic means, colour coding, graphical timelines or similar display features).
- Having different views of the same data is a useful feature, because it allows the clinician to spot correlation between events that they may have missed otherwise.
- Summaries that provide a 30-second overview of the patient's history are often desireable; ideally, these should fit entirely on a computer screen. However, users should be able to obtain more detailed information about specific events by expanding their description.

Following these requirements, we proposed in this project an integrated visualisation tool where users can use a graphical interface coupled with a text generation engine in order to navigate through patient records. Textual reports have the advantage of offering a snaphsot view of a patient's history at any point in time, they can be used for checking the consistency of a patient's record, can be ammended and printed, used in communication between clinicians or clinicians and patients. Text is a good way of describing temporal information (events that happened at a certain position in time with respect to another event), of summarising numerical data (for example, specifying that *liver tests were normal* instead of listing individual measurements for billirubin concentration, Alanine aminotransferase, Alkaline phosphatase, Aspartate aminotransferase, albumin and total protein). However, pure text is not always the best medium of presenting large amount of information, part of which is numerical and most of which is highly interconnected. Text loses the ability of navigating through information, of expanding some events and of highlighting important dependencies between pieces of information. A textual report alone cannot effectively combine the time sequence element with the semantic dependencies - both of which are essential in representing patient records. Depending on the type of report chosen, either one or the other of these elements will necessarily be emphasised at the expense of the other.

We envisage therefore that, depending on circumstances, users may want to have fully textual reports (for example, for producing printed summaries of a patient's history) or combined graphical and textual reports (for interactive visualisation). In the following, we will describe the two reports generated in either of the two scenarios. Section 3 will describe in more detail the natural language generation techniques employed in generating both independent textual reports and report snippets that support the graphical interface.

2.1 Textual reports

Textual reports are views of a data-encoded electronic patient record (a chronicle), which is itself both a distillation and an integration of the elements within the traditional EPR. In this respect, they do not correspond to the narratives traditionally contained in a patient record, such as letters from clinicians, discharge notes, consult summaries. They are a new type of text that aggregates information from the full record.

Based on our requirements analysis with clinicians, we identified two main types of textual report that could be used in different settings. The first is a longitudinal report, which is meant to provide a quick historical overview of the patient's illness, whilst preserving the main events, such as diagnoses, investigations and interventions. It describes the events in the patient's history ordered chronologically and grouped according to the type. It contains most events in the history, although some preliminary filtering is performed to remove usually a small number of isolated events. The following example displays a fragment of a generated longitudinal summary.

```
(1)  The patient is diagnosed with grade 9
invasive medullary carcinoma of the breast.
She was 39 years old when the first malignant
cell was recorded.  The history covers 1517
```

```
weeks, from week 180 to week 1697. During this
time, the patient attended 38 consults.
```

YEAR 3:

Week 183

- Radical mastectomy on the breast was performed to treat primary cancer of the left breast.
- Histopathology revealed primary cancer of the left breast.

Week 191

- Examination revealed no enlargement of the liver or of the spleen, no lymphadenopathy of the left axillary lymphnodes,no abnormality of the haemoglobin concentration or of the leucocyte count.
- Radiotherapy was initiated to treat primary cancer of the left breast.
- ...

The second class of summary focuses on a given type of event in a patient's history, such as the history of diagnoses, interventions, investigations or drug prescription. In contrast to the longitudinal summaries, which are generic, this type of report is query-oriented, since it summarizes only events which the user deems relevant.

A summary of the diagnoses, for example, will focus on the `Problem` events that are recorded in the chronicle, whilst other events only appear if they are directly related to a `Problem`. This type of summary is necessarily more concise, since the events do not have to appear chronologically and thus can be grouped in larger clusters. Secondary events are also more highly aggregated. For example:

```
(2)  In week 483, histopathology revealed
primary cancer of the right breast.  Radical
mastectomy on the breast was performed to treat
the cancer.
In week 491, no abnormality of the leucocyte
count or of the haemoglobin concentration,
no lymphadenopathy of the right axillary
lymphnodes, no enlargement of the spleen or
of the liver and no recurrent cancer of the
right breast were revealed.  Radiotherapy was
initiated to treat primary cancer of the right
breast.
In the weeks 492 to 496, five radiotherapy
cycles were performed.
```

A subclass of reports in this category is represented by reports of selective events. For example, a clinician may suspect that a certain patient has interrupted their chemotherapy package repeatedly and wants to see if this correlates with a certain medical condition such as anaemia or if there are other causes behind it. In this case, they may order a report focused on incomplete chemotherapy packages and investigations of type blood test.

2.2 Visual reports

A visual report is a one-screen overview of a patient record (see fig.1), where various types of events are colour-coded and displayed along a timeline. Selection of events can be used for highlighting event dependencies or for generating focused textual reports. Apart from general history timelines, users can also investigate the trend of numerical values, for example increases and decreases in the billirubin concentration from one test to an other.

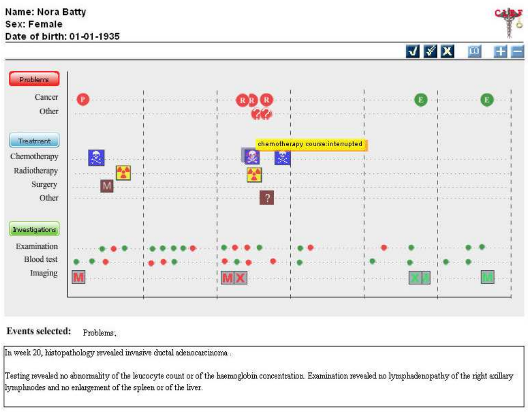

Figure 1: Visual history snapshot: minimum zooming, the user has selected *Problems* in the graphical interface and a summary of all recorded problems is displayed in the bottom pane.

The advantage of a visual display is that the user can have a global view of a patient's history. However, much information is hidden behind each event displayed on the timeline. The user can reveal this information by interacting with the graphical display. By zooming in or out, events are collapsed or expanded. For example, in Fig.1 there is a chemotherapy event spanning 8 weeks. In a minimum zoom view, they appear as a single chemotherapy event; by zooming in, the user will be able to see that there have been 6 chemotherapy cycles given succesfully, and 3 chemotherapy cycles have been deferred. Hovering the mouse over any event will display as a tooltip a short description of the event. For example, hovering over the chemotherapy event in Fig. 1, the user will see the tooltip *A complete chemotherapy course was given from week 312 to week 320*. Further information about an event can be obtained by clicking on its icon. A chemotherapy event, for example, "hides" information about the particular drug regimen used, exact dates of chemotherapy cycles and reasons for deferring a particular cycle. Since this information is better expressed as text than graphically, each selection of an event will trigger the production of a report snippet that describes in more detail that particular event.

Apart from individual events, the user can also select multiple events (by clicking on several event icons on the timeline), classes of events (by clicking on the event name on the left hand side of the screen) or time spans (by selecting years on the horizontal axis). The effect of such selections will be the production of summaries similar to those described in the previous section. Selection of events will produce event-focused summaries, whilst selection of time spans will produce longitudinal summaries for that particular span.

Semantic relations between events are displayed on demand, allowing the user to see the logical sequence of events (tracing, for example, the reason for performing a red packed cell transfusion to anaemia which was in turn caused by chemotherapy performed to treat cancer).

3 The Report Generator

In the following, the term *Report Generator* will be used to designate the software that performs text generation, as a result of either a direct request from the user for a specific type of report or a selection of events in the graphical timeline. The output of the report generator may be either a full report or a report snippet, but practically, the type of selection employed in chosing the focus of the report does not influence the technique used in generating it.

3.1 Input

The input to the Report Generator is a chronicle, which is a highly structured representation of an Electronic Patient Record, in the form of a semantic network. It is beyond the scope of this paper to describe the methodology involved in transforming an EPR into a chronicle - the chronicalisation process is complex and involves Information Extraction from narratives, solving multi-document coreference, temporal abstraction and inferencing over both structured and information extraction data (Harkema et al., 2005). For the purpose of this paper, we consider the input correct and complete. The main advantage in using a chronicle as opposed to a less structured Electronic Patient Record lies in the richness of information provided. Having access to not only facts, but to the relations between them, has important implications in the design of the content selection and text structuring stages. This facilitates better and easier text generation and allows for a higher degree of flexibility of the generated text.

The chronicle relations can be categorised into three types according to their role in the generation process. *Rhetorical relations* are relations of causality and consequence between two facts (such as, `Problem` CAUSED-BY `Intervention` or `Intervention` INDICATED-BY `Problem`) and are used in the document planning stage for automatically inferring the rhetorical structure of the text, as it will be described in 3.2.2. *Ontological relations* such as `Intervention` IS-SUBPART-OF `Intervention` bear no significance in text planning and realisation, but can be used in content selection. *Attribute relations* such as `Problem` HAS-LOCUS `Locus` or `Investigation` HAS-INDICATION `Problem` are used in grouping messages in a coherent way, prior to the construction of the rhetorical structure tree.

3.2 System design

The system design of the Report Generator follows a classical NLG pipeline architecture, with a Content Selector, Content Planner and Syntactic Realiser. The Content Planner is tightly coupled with the Content Selector, since part of the document structure is already decided in the event selection phase. Aggregation is mostly conceptual rather than syntactic, therefore it is performed in the content planning stage as well.

3.2.1 Content selection

The process of content selection is driven by two parameters: the type of summary and the length of summary. We define the concept of *summary spine* to represent a list of concepts that are essential to the building of the summary. For example, in a summary of the diagnoses, all events of type `Problem` will be part of the spine (Figure 2). Events linked to the spine through some kind of relation may or may not be included in the summary, depending on the specified type and length of the summary. The design of the system does not restrict the spine to containing only events of the same type: a spine may contain, for example, `Problems` of type *cancer*, `Investigations` of type *x-ray* and `Interventions` of type *surgery*.

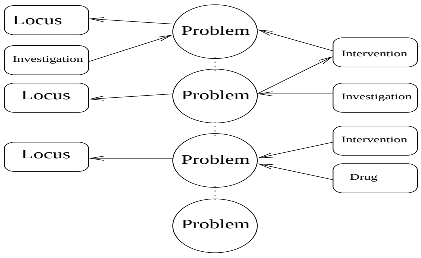

Figure 2: Spine events for a summary of diagnoses

The relations stored in the chronicle help in the construction of clusters of related events. A typical cluster may represent, for example, that a patient diagnosed with cancer following a clinical examination, a mastectomy was performed to remove the tumour, a histopathological investigation on the removed tumour confirmed the cancer, radiotherapy was given to treat the cancer, which caused an ulcer that was then treated with some drug. Smaller clusters are generally not related to the main thread of events, therefore the first step in the summarisation process is to remove small clusters[2] The next step is the selection of important events, as defined by the type of summary. Each cluster of events is a strongly connected graph, with some nodes representing spine events. For each cluster, the spine events are selected, together with all nodes that are at a distance of less than n from spine events, where n is a user-defined parameter used to adjust the size of the summary. For example, in the cluster presented in figure 3, assuming a depth value of 1, the content selector will choose *cancer*, *left breast* and *radiotherapy* but not *radiotherapy cycle*, nor *ulcer*.

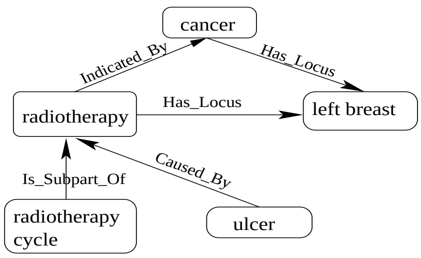

Figure 3: Example of cluster

The result of the content selection phase is a list of messages, each describing an event with some of its attributes specified. The number of attributes specified depends on the depth level of a message (i.e., how far from the spine the event is). For example, a `Problem` event has a large number of attributes, consisting of `name, status, existence, number of nodes counted, number of nodes involved, clinical course, tumour size, genotype, grade, tumour marker` and `histology`, along with the usual time stamp. If the *Problem* is a spine event, all these attribues will be specified, whilst if the *Problem* is 2 levels away from the spine, only the name and the existence will be specified.

3.2.2 Document planning

The document planner component is concerned with the construction of complete document plans, according to the type of summary and cohesive relations identified in the previous stage. The construction of document plans is, however, initiated in the content selection phase: content is selected according to the relations between events, which in turn informs the structure of the target text.

The document planner uses a combination of schemas and bottom-up approach. A report is typically formed of three parts: a schematic description of the patient's demographic information (e.g., name, age, gender); a two-sentence summary of the patient's record (presenting the time span of the illness, the number of consults the patient attended and the number of investigations and interventions performed); and the actual summary of the record produced from the events selected to be part of the content. In what follows, we will concentrate on this latter part.

The first stage in structuring the summary is to combine messages linked through attributive relations. This results in instances such as that shown in example (3), where a `Problem` message is combined with a `Locus` message to give rise to the composite message `Problem-Locus`.

In the second stage, messages are grouped according to specific rules, depending on the type of summary. For longitudinal summaries, the grouping rules will, for example, stipulate that events occurring within the same week should be grouped together, and further grouped into years. In event-specific summaries, patterns of similar events are first identified and then grouped according to the week(s) they occur in; for example, if in week 1 the patient was examined for enlargement of the liver and of the spleen with negative results and in week 2 the patient was again examined with the same results and

[2] In the current implementation these are defined as clusters containing at most three events. This threshold was set following empirical evidence.

underwent a mastectomy, two groups of events will be constructed, leading to output such as:

```
(3)  In weeks 1 and 2, examination of the
abdomen revealed no enlargement of the liver or
of the spleen.
In week 2, the patient underwent a mastectomy.
```

Within groups, messages are structured according to discourse relations that are either retrieved from the input database or automatically deduced by applying domain specific rules. At the moment, the input provides three types of rhetorical relation: Cause, Result and Sequence. The domain specific rules specify the ordering of messages, and always introduce a Sequence relation. An example of such a rule is that a histopathology event has to follow a biopsy event, if both of them are present and they start and end at the same time. These rules help building a partial rhetorical structure tree. Messages that are not connected in the tree are by default assumed to be in a List relation to other messages in the group, and their position is set arbitrarily. Such events are grouped together according to their type; for example all unconnected `Intervention` events, followed by all `Investigations`.

In producing multiple reports on the same patient from different perspectives, or of different types, we operate under the strong assumption that event-focussed reports should be organised in a way that emphasises the importance of the event in focus. From a document structure viewpoint, this equates to building rhetorical structures where the focus event (i.e., the spine event) is expressed in a nuclear unit, and skeleton events are preferably in sattelite units.

At the sentence level, spine events are assigned salient syntactic roles that allows them to be kept in focus. For example, a relation such as `Problem` CAUSED-BY `Intervention` is more likely to be expressed as:

> “The patient developed a `Problem` as a result of an `Intervention`.”

when the focus is on `Problem` events, but as:

> “An `Intervention` caused a `Problem`.”

when the focus is on `Interventions`.

This kind of variation reflects the different emphasis that is placed on spine events, although the wording in the actual report may be different. Rhetorical relations holding between simple event descriptions are most often realised as a single sentence (as in the examples above). Complex individual events are realised in individual clauses or sentences which are connected to other accompanying events through the appropriate rhetorical relation. Additionally, the number of attributes included in the description of a `Problem` is a decisive factor in realising an event as a phrase, a sentence or a group of sentences. In the following two examples, there are two `Problem` events (`cancer` and `lymphnode count`) linked through an `Investigation` event (`excision biopsy`), which is indicated by the first problem and has as a finding the second problem. In Example 4, the problems are first-mentioned spine events, while in Example 5, the problems are skeleton events (the cancer is a subsequent mention and the lymphnode count is a first mention), with the `Investigation` being the spine event.

```
(4)  A 10mm, EGFR +ve, HER-2/neu +ve, oestrogen
receptor positive cancer was found in the
left breast (histology:  invasive tubular
adenocarcinoma).  Consequently, an excision
biopsy was performed which revealed no
metastatic involvement in the five nodes
sampled.
(5)  An excision biopsy on the left breast
was performed because of cancer.  It revealed
no metastatic involvement in the five nodes
sampled.
```

As these examples show, the same basic rhetorical structure consisting of three leaf-nodes and two relations (CAUSE and CONSEQUENCE) is realised differently in a `Problem`-focussed report compared to an *Investigation*-based report. The conceptual reformulation is guided by the type of report, which in turn has consequences at the syntactic level.

3.2.3 Aggregation

The fluency of the generated text is enhanced by conceptual aggregation, performed on messages that share common properties. Simple aggregation rules state, for example, that two investigations with the same name and two different target loci can be collapsed into one investigation with two target loci. Consider, for example, a case where each clinical examination consists of examinations of the abdomen for enlargement of internal organs (liver and spleen) and examination of the lymphnodes. Thus, each clinical examination will typically consist of three independent Investigation events. If fully expanded, a description of the clinical examination may look like:

(6) • examination of the abdomen revealed no enlargement of the spleen
• examination of the abdomen revealed no enlargement of the liver
• examination of the axillary lymphnodes revealed no lymphadenopathy of the axillary nodes

With a first level of aggregation, this is reduced to:

(7) Examination revealed no enlargement of the spleen or of the liver and no lymphadenopathy of the axillary nodes.

However, even this last level of aggregation may be not enough, since clinical examinations are performed repeatedly and consist of the same types of investigation. We employ two strategies for further aggregating similar events. The first solution is to report only those events that deviate from the norm - for example, abnormal test results. The second, which leads to larger summaries, is to produce synthesised descriptions of events. In the case of clinical examinations for example, it can describe a sequence of investigations such as the one in Example 7 as *"The results of a clinical examination were normal"*, or, if the examination result deviates from the norm on a restricted numbers of parameters, as *"The results of clinical examination were normal, apart from an enlargement of the spleen"*.

4 Related work

Natural language generation has been used in the medical domain for various applications. For example: to generate drug leaflets (i.e., pill inserts) in multiple languages and styles (PILLS (Bouayad-Agha et al., 2002)), letters to patients to help them stop smoking (STOP (Reiter et al., 2003)), individualised patient-education brochures (MIGRANE (Buchanan et al., 1992)); HealthDoc (Hirst et al., 1997)); Piglit (Binsted et al., 1995)). There is also a body of work on the generation of summaries of patient records (e.g., (Afantenos et al., 2005), (Elhadad and McKeown, 2001)). This work, however, differs from ours in that they concentrate on the summarization of textual records, while we deal with summarization of data from Electronic Patient Records.

Most computer-based patient record management systems have simple generation facilities built-in, which produce simple text, normally consisting of unconnected sentences and thus lacking fluency. Natural language generation techniques have been applied in various reporting systems for generating telegraphic textual progress notes (Campbell et al., 1993), reports on radiographs (A. Abella, 1995), and bone scans (Bernauer et al., 1991) or post-operative briefings (M. Dalal, 1996).

The timeline method has been used extensively in visualising patient histories. The Lifelines project (Plaisant et al., 1998) provides a method for visualising and accessing personal histories by means of a graphical interface, and has been used for both patient records and legal case histories. TeleMed (Kilman and Forslund, 1997) gathers patient information from distributed databases and presents it in a Web interface as icons on a timeline; interaction with the icons provides access to more detailed descriptions of the individual pieces of information. Various authors describe the advantages of the timeline approach to visualising temporal data of the kind present in patient histories (Tufte and Kahn, 1983; Cousins and Kahn, 1991).

5 Conclusion

We presented in this paper an innovative approach to the problem of presenting and navigating through patient histories as a means of supporting clinical care and research. Our approach uses a visual naviagation tool combined with natural language generated reports. Although developed for the domain of cancer, the very same methods that we could be used with little adaptation effort for general health care; they are based on a general model of the main events in a patient's medical history that occur across all diseases and ailments: symptoms, diagnoses, investigations, treatments, side effects and outcomes. As such, we only make use of general medical knowledge, which applies to any medical sub-domain. The design of both the text generator and visual navigator is completely data-driven by the system input, the chronicle.

An important consequence of the use of chronicles as input is that our system does not require complex domain semantics, which has been regarded as one of the essential components of NLG systems (Reiter and Dale, 2000). This is partly because inferences that are normally required to combine and order facts in the generated summary have already been performed prior to the language generation process, and their results have been stored in the chronicle as relations between facts. Indeed, a key feature of our system is that — apart from the relations present in the input data — it does not use any kind of external domain knowledge in the process of content selection. The only domain specific rules used are in text organisation, specifically in the ordering of messages; these are not essential, although they do improve the fluency of the text. Our lack of reliance on domain semantics is a clear advantage for the portability of the system to other domains. It is nevertheless true that more specific domain knowledge could improve the summarization process, for example, in deciding which events should be considered important (which will clearly vary from one medical area to the next). In our report generation system, this type of knowledge can be encoded as a set of external rules, whose application would not be essential to the system. These rules can be specified without interfering with the main application, and require no changes in previous code.

Current electronic patient records are human-friendly but highly impoverished from the point of view of systematic machine analysis and aggregation. By contrast, the ideal machine representation is far too complex to be human-friendly. Our research suggests that with the combined graphical and natural language genereration approach we have described here, this complex machine representation can be made both relatively familiar, and friendly.

Acknowledgment

The work described in this paper is part of the Clinical E-Science Framework (CLEF) project, funded by the Medical Research Council grant G0100852 under the E-Science Initiative. We gratefully acknowledge the contribution of our clinical collaborators at the Royal Marsden and Royal Free hospitals, colleagues at the National Cancer Research Institute (NCRI) and NTRAC and to the CLEF industrial collaborators.

References

J. Starren A. Abella, J. Kender. 1995. Description generation of abnormal densities found in radiographs. In *Proceedings of the Symposium on Computer Applications in Medical Care (SCAMC)*, pages 542–546.

Stergos D. Afantenos, Vangelis Karkaletsis, and Panagiotis Stamatopoulos. 2005. Summarization from medical documents: A survey. *Artificial Intelligence in Medicine*, 33(2):157–177.

J. Bernauer, K. Gumrich, S. Kutz, P. Linder, and D.P. Pretschner. 1991. An interactive report generator for bone scan studies. In *Proceedings of the Symposium on Computer Applications in Medical Care (SCAMC)*, pages 858–860.

K. Binsted, A. Cawsey, and R.B. Jones. 1995. Generating personalised patient information using the medical record. In *Proceedings of Artificial Intelligence in Medicine Europe*, Pavia, Italy.

N. Bouayad-Agha, R. Power, D. Scott, and A. Belz. 2002. Pills: Multilingual generation of medical inf ormation documents with overlapping content. In *Proceedings of LREC 2002*, pages 2111–2114.

B.G. Buchanan, J.D. Moore, D.E. Forsythe, G. Carenini, and S. Ohlsson. 1992. Involving patients in health care: explanation in the clinical setting. In *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care (SCAMC'92)*, pages 510–512.

K.E. Campbell, K. Wieckert, L.M. Fagan, and M.Musen. 1993. A computer-based tool for generation of progress notes. In *Proceedings of the Symposium on Computer Applications in Medical Care (SCAMC)*, pages 284–288.

Cousins and M. Kahn. 1991. The visual display of temporal information. In *Artificial Intelligence in Medicine*, pages 341–357.

N. Elhadad and K. McKeown. 2001. Towards generating patient specific summaries of medical articles. In *Proceedings of the Workshop on Automatic Summarization, NAACL 2001*, Pittsburg, USA.

H. Harkema, I. Roberts, R. Gaizauskas, and M. Hepple. 2005. Information extraction from clinical records. In *Proceedings of the 4th UK e-Science All Hands Meeting*, Nottingham, UK.

Graeme Hirst, Chrysanne DiMarco, Eduard H. Hovy, and K. Parsons. 1997. Authoring and generating health-education documents that are tailored to the needs of the individual patient. In *Proceedings of the 6th International Conference on User Modeling*, Italy.

David G. Kilman and David W. Forslund. 1997. An international collaboratory based on virtual patient records. *Communications of the ACM*, 40(8):110–117.

K.McKeown M. Dalal, S. Feiner. 1996. MAGIC: An experimental system for generating multimedia briefings about post-bypass patient status. *Journal of the American Medical Informatics Association*, pages 684–688.

C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman. 1998. Lifelines: Using visualization to enhance navigation and analysis of patient records. In *Proceedings of the 1998 American Medical Informatic Association Annual Fall Symposium*, pages 76–80.

Alan Rector, Jeremy Rogers, Adel Taweel, David Ingram, Dipak Kalra, Jo Milan, Robert Gaizauskas, Mark Hepple, Donia Scott, and Richard Power. 2003. Clef - joining up healthcare with clinical and post-genomic research. In *Second UK E-Science "All Hands Meeting"*, Nottingham, UK.

Ehud Reiter and Robert Dale. 2000. *Building natural language generation systems*. Cambridge University Press, New York, NY, USA.

E. Reiter, R. Robertson, and L.M. Osman. 2003. Lessons from a failure: Generating tailored smoking cessation letters. *Artificial Intelligence*, 144:41–58.

E.R. Tufte and M. Kahn. 1983. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut.