Title: "A Physician's Authoring Tool for Generation of Personalized Health Education in Reconstructive Surgery

Chrysanne Di Marco,[1] Don Cowan, [1] Peter Bray,[2] Dominic Covvey, [1] Vic Di Ciccio, [1] Eduard Hovy,[3] Joan Lipa,[2] Doug Mulholland[1]

[1] University of Waterloo

Waterloo, Ontario Canada

cdimarco@uwaterloo.ca

[2] Division of Plastic Surgery, Department of Surgery and Department of Surgical Oncology,

University Health Network, University of Toronto

Toronto, Ontario, Canada

[3] Information Sciences Institute

University of Southern California

USA

Abstract

Personalization of health information related to a specific patient's characteristics would be an effective means of providing patient-centric health care. Further, it would allow the patient to be both better-educated about their specific condition and better able to make informed decisions. This paper contains a description of a method of providing such personalized health-related materials in one area of plastic surgery. The paper contains an outline of the approach, a description of the tools developed to date, and a design for an authoring tool that is being created to support the domain expert in acquiring and organizing health information.

Introduction.

E-health services are playing an increasingly important role in health-care management by providing relevant and timely information to patients about their medical care. An important factor in the rapid growth of online health services is the trend in health management to patient-centric health care: patient-centric care aims to involve the patient directly in the medical decision-making process by providing better access to the relevant information that patients need to understand their medical condition and to enable them to make more-informed decisions about their prescribed treatment.

An effective means of providing patient-centric health care would be through the personalization of health information: with individually tailored information, the patient would be both better-educated about their specific condition and better able to make informed decisions. However, health-information material is often limited in its effectiveness by the need to address it to a wide audience.

What is generally produced is either a minimal, generic document that contains only the information common to everyone, or a maximal document that tries to provide all the information that might be relevant to someone (and hence much that is irrelevant to many). In either case, the document is likely to be ineffective in patient education, with a resulting low level of patient compliance with prescribed treatments. Recognizing this, health educators have paid much attention to methods of identifying different segments of their audience and their differing needs and constructing material accordingly, but at some level, the material remains generic.

The goal of our research is to develop Natural Language Generation methodologies and software tools that can be used to deliver health educational materials tailored to the needs of the individual patient in a timely and accessible manner through Web-based systems. Our initial domain of application is reconstructive breast surgery, but the Natural Language techniques and tools we are developing will be generally applicable to all medical interventions.

Authoring as a Means of Producing Tailored Documents.

The original HealthDoc Project (1994-1999) (DiMarco et al. 1995; Hirst et al.1997) produced three authoring tools for creating tailorable documents: Jakeway's (Jakeway 1995) SPLAT authoring tool created "Master Documents" (i.e., a representation of all the content that might ever be needed for any user) at a deep-syntactic level of representation; Parsons' (Parsons 1997) authoring tool allowed the health-education writer to directly author the master document, including marking up the document with rhetorical and other linguistic information; Banks (Banks 1999) developed an authoring tool which integrated the two previous approaches by automatically transforming a surface-text master document into the underlying deep-syntactic representation which could be the input to a generation system.

The creation of the input representation for Natural Language Generation systems is a problem for all such systems. The issue of where we obtain the information which specifies the actual text to be generated is one of the two major problems intrinsic to Natural Language Generation. The approach we are taking -- `preparing' a database of input specifications by authoring material for later use in generating new documents crafted from the pre-existing material -- is a promising and feasible methodology for reducing the complexities inherent in trying to generate high-quality, stylistically expressive text `from scratch'.

The Need for Tailored Patient Education in Reconstructive Breast Surgery.

We have chosen to focus on personalization of patient education in the domain of reconstructive breast surgery as a representative example of the kinds of difficult problems that need to be addressed in developing automated Natural Language tailoring systems. Interventions in modern surgical oncology are often complex, multi-step procedures involving multiple surgeons or surgical subspecialties. In

considering post-ablative reconstruction alone, a variety of surgical options may be available in a given situation. Each of these options will be associated with different advantages, disadvantages, and peri-operative implications. It is a challenge for both patients and surgeons to ensure that sufficient information has been communicated preoperatively about these procedures.

Although preoperative information brochures have documented value for patient education, a library of static documents would be difficult to establish if it were to encompass all reconstructive surgical alternatives. For a patient undergoing a multi-step procedure, a handful of brochures would be required, which would lack cohesiveness and would likely be very confusing. Consequently, existing preoperative information brochures are only available for the most common reconstructive surgical procedures and must, by necessity, remain generic in nature to ensure applicability to all patients.

Although no amount of supplemental documentation can replace the surgeon-patient dialogue with which informed consent is obtained, it is well-documented that only a small fraction of the information communicated in this process is actually retained by the patient. Reference material for review by patient, friends, and significant others would have great value in the pre-operative, peri-operative and postoperative stages if this information could be tailored to the individual patient. However, the complexity of the surgical procedure and the variety of options that need to be considered in tailoring documentation to the individual patient make the creation of appropriate material a combinatorially explosive process.

In previous work in the HealthDoc Project (Jakeway 1995; DiMarco, Hovy, and Parsons 1997; Parsons 1997; Banks 1999; Banks and DiMarco 2000), DiMarco and her students developed several authoring tools for the creation of tailorable content in the HealthDoc master-document format. However, none of these tools was geared to the domain expert but rather were intended for a computer programmer or computational linguist.

Up until now, authoring of patient-education materials has typically been accomplished through the interaction of the health professional with a `knowledge engineer', someone trained in structured knowledge acquisition. We are developing a supportive authoring environment that will allow the domain expert -- physicians, surgeons, and other health-care providers -- to interact directly with the Natural Language tailoring system to enter the textual variants that will later be used to produce the tailored versions. This involves two levels of authoring: definition of the discourse structure of the master document and creation of the actual text content:

At the knowledge level, the physician-author will be guided through an interactive dialogue to identify the features that will be used to create the high-level discourse structure for the Master Document (DiMarco, C., Foster, M.E. 2005). Yang (Yang 2005) developed the framework for a knowledge acquisition process that would handle the knowledge level of modeling in our authoring tool.) At the content level of authoring, the core of the authoring tool

will provide facilities for entering text variants for any domain, i.e., not just patient education. Each text variant will be annotated with a Boolean condition made up of a combination of relevant reader features. Additional mark-up on pieces of the master document may include layout and other formatting details. Authoring features will be provided for: gradual editing of a master document (allowing editing at any re-entry point in the document); checking conditions on text variants for consistency and completeness; displaying sets of text variants for evaluation and further editing.

Additional features will be provided in the full authoring environment to guide a physician-author through the transformation of the informal information they convey to patients into a formal master document. This authoring environment will assist the surgeon by guiding the mapping from each stage in the surgical procedure to the relevant parts of the master document that need to be edited. This will involve: constructing a model of the surgical procedure to allow checking for consistency and completeness of the master-document content; allowing adaptation of the model according to the preferences of multiple physician-authors; providing different views of the master-document content according to corresponding views of the surgical procedure.

The HealthDoc Authoring Tool: Current Status

We have now designed an XML-based notation for our original tailorable document ("Master Document") representation, which included significant extensions to the original specification of a Master Document. A set of XSL transformations has been developed that automatically convert Master Documents to "Dynamic Documents." Dynamic documents from a user perspective are Web-based documents which accept as input responses to a series of questions and conditions and produce a customized version of the Master Document according to the inputs. We subsequently designed and implemented a prototype Web-based document tailoring system to produce Dynamic Documents from the underlying Master Document. This prototype tailoring system has so far been tested on five sample Master Documents.

We have also developed an initial corpus of tailored educational content for reconstructive breast surgery that will form the basis of the text database of patient education material for our project. An initial "pre-authoring tool" allows the Excel representation of the tailored content that our surgeon team members have been developing to be automatically converted into the new XML-based Master Document, and from thence to a Web-based Dynamic Document from which tailored versions of the content may be generated.

The Benefits of XML Tagging in the Master Document

We have chosen XML-based tagging for the Master Document because such a representation is highly extensible in many ways. There are many tools based on XSL transformations or cascading style sheets (CSS) that

can transform the entire tagged document into alternate representations that add value to the original.

The Master Document contains conditions that determine the choice of text fragments to be displayed and so an intermediate representation as a program with its conditional capabilities is a good option. We have implemented an XSL transformation that generates a PHP program that contains all the text of the Master Document with conditional statements corresponding to the conditions in the Master Document. The XSL transformation or compiler can be customized in many ways including "run-time" document analysis and debugging. Error diagnostics, which can be built into the transformed Master Document, can be easily referenced back to the original Master Document.

The output produced by the PHP program can itself be XML, thus allowing further customization through XSL or CSS transformations. Thus the document could be retargeted for display in many different contexts such as computer displays, printers, mobile browsers on cell phones, PDAs or Blackberries.

Another XSL transformation has been implemented to create the customization form from the Master Document. When users have filled in the form, a button can be pressed to present the dynamic document, customized to the user's selections from the form.

The author(s) of the Master Document can use a step-wise refinement process to build the document. They can easily build an outline and then refine individual chapters, sections or smaller elements, right down to individual lexical elements such as words, characters or images. At any time they can invoke the XSL transformation and view the results of their efforts. This type of operation works well as a section of the document pertaining to a specific medical condition or surgical procedure is refined.

XML tagging supports extra annotations, such as stylistic attributes, which can be specified for every level of document element such as an entire document, a chapter, or even a paragraph. This format allows us to support such things as a default layout, font or character style or elements within an enclosing element that can be assigned attributes that override the defaults.

Recursive document definitions and references are permitted. A document section can easily incorporate another document, elements of other documents or documents and sections that are already in the active document element tree.

XML documents can be transformed into many different forms. The document-object model (DOM) and associated tools could be used to represent the document and invoke other types of processing. The documents could also be transformed into formal notations such as logic or graphical structures. Thus one could ask questions about reachability of document sections or precedence.

There are document management tools that support functions such as check-in and check-out thus allowing multiple users to work on different parts of a document at the same time.

Our research project has access to the Web-based Informatics Development Environment (WIDE) Toolkit

(Cowan, D., Fenton, S., and Mulholland, D. 2005) for creating web-based systems. WIDE uses an underlying declarative model based on XML. With WIDE, the developer is prompted via a "wizard" to specify the various components needed to develop a web-based system that can support databases and rich multimedia content. The developer can also specify access controls so that groups can take ownership of their own data thereby distributing the problem of data creation and maintenance. Further, the access controls support the concept of a shared virtual space where users are able not only to view information, but also to augment it. Extensive WIDE document management tools permit changes to be logged and both source and production versions (i.e., personalized versions) of documents can be indexed, searched, cross-referenced and spell-checked.

The Next Generation Authoring Tool

We have designed and implemented a number of tools to process and display the tagged documents after they are created. However, we need to produce tools that will support the domain expert as they produce the text fragments that will eventually form a document to be customized for a specific patient. In this section we present our thinking on the design of the authoring tool that the medical domain expert will use while creating Master Documents.

The documents that are being constructed by the domain experts are quite complex. At a minimum they consist of many text fragments, where there can be many alternatives for each text fragment. In a simple example, patients undergoing a specific surgical procedure might expect to see quite different outcomes depending upon their age group. Thus, an outcomes paragraph or section could have many different associated outcome-based explanations.

Producing such a document, ensuring that each conditional section is complete and accurate is intellectually demanding and requires careful organization. An editor that supports this activity should have the following characteristics:

The editor should be able to position itself easily over the text fragment that is being created or edited and show all the alternatives

The editor should operate in a "what you see is almost what you get" mode. At any time the domain expert should be able to insert parameters that define a patient type and the document should be able to be viewed within the editor. One can view the editor as a complex word processor with windows defined by a domain expert that can outline the scope of text seen by that expert.

The editor should support the domain expert in that the editor environment should be complete enough that domain experts do not have to use external tools to see fragments and their alternatives or to view a complete parameterized view of the document.

Other design features of the document authoring tool will come to light as we complete the design and implementation of the entire system for creating the Master Document within the current restricted domain. We have currently acquired a base tool that we believe can be

customized to provide all the functionality required for the domain experts.

Summary

In this paper we have presented a description of a document structure and an environment for producing personalized health education documents that are tailored to the characteristics and requirements of the individual. Personalization can take many forms based on parameters such as age, gender, language, and literacy level.

The environment has two basic functions; the creation of the document and its final production for a client with specific personal characteristics. We have concluded with a description of the next phase of our research which is an authoring tool designed to support the domain expert in creating a Master Document. In this regard we have described the authoring tool's major characteristics.

Acknowledgements.

Aspects of the research described in this paper were funded by Bell Canada, Communications and Information Technology Ontario (CITO), Human Resources Development Canada (HRDC), iAnywhere Solutions (Sybase), IBM, Natural Sciences and Engineering Research Council of Canada (NSERC), Rogers Cable, and Siebel Systems.

References

Banks, S. 1999. English to sentence plan translation using subsumption on stylistic abstraction. M.Math.diss., Dept of Computer Science, Univ. of Waterloo.

Banks, S., and DiMarco, C. 2000. Surface form to sentence plans: A method for

English-to-English translation. In *Proceedings of the 4th Symposium on Natural Language Processing*.

Cowan, D.; Fenton, S.; and Mulholland, D. 2005. Web-based Informatics Development Environment (WIDE) Toolkit - http://csg.uwaterloo.ca/wide.htm

DiMarco, C., and Foster, M.E. 2005. A method and apparatus for authoring of customizable multimedia documents. International patent application PCT CA98/00771 filed 11 August 1998 (priority dating from U.K. application number 9716986.6 filed 12 March 1997). U.S. patent granted March 2005. Canadian patent pending.

DiMarco, C.; Hirst, G., and Hovy, E. 1997. Generation by selection and repair as a method for adapting text for the individual reader. Workshop on Flexible Hypertext, Eighth ACM International Hypertext Conference, ACM Hypertext Conference.

DiMarco, C.; Hirst, G.; Wanner, L., and Wilkinson, J. 1995. HealthDoc: Customizing patient information and health education by medical condition and personal characteristics. Workshop on Artificial Intelligence in Patient Education, University of Glasgow.

Hirst, G.; DiMarco, C.; Hovy, E.; and Parsons, K. 1997. Authoring and generating health-education documents that are tailored to the needs of the individual patient. In *Proceedings of the Sixth International Conference on User Modeling*, Sardinia, Italy.

Jakeway, B. 1995. SPLAT: A sentence plan authoring tool for natural language

Generation, M.Math.diss. Dept of Computer Science, University of Waterloo.

Parsons, K. 1997. An authoring tool for customizable documents. M.Math.diss. Dept of Computer Science University of Waterloo.

Yang, C. 2005. Design of a knowledge acquisition tool using a constructivist approach for creating tailorable patient education materials. MMath.diss. School of Computer Science, University of Waterloo.