The Development of a Natural Language Generation System For Personalized e-Health Information

C. DiMarco[a], H. Dominic Covvey[b], P. Bray[c], D. Cowan[a,d], V. DiCiccio[a], E. Hovy[e], J. Lipa[c], D. Mulholland[d]

[a]*David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada*
[b]*Waterloo Institute for Health Informatics Research, University of Waterloo, Waterloo, Ontario, Canada*
[c]*Division of Plastic Surgery, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada*
[d]*Conputer Systems Group, University of Waterloo, Waterloo, Ontario, Canada*
[e]*Information Sciences Institute, University of Southern California, USA*

Abstract

Personalization—adaptivity to the individual—is becoming an essential component of any computer-based system. In e-health systems, personalization of health information is emerging as a key factor in the trend to *patient-centric* care. Patient-centric healthcare aims to *engage* patients in their treatment to promote greater compliance and satisfaction with their therapeutic regimens, resulting in both better patient outcomes and reduced healthcare costs. We have developed a prototype Web-based Natural Language Generation system for the authoring and subsequent personalization of patient education materials. Our initial domain of application is reconstructive breast surgery, but our Natural Language software tools and authoring methodologies are generally applicable to all medical interventions.

Keywords:

Patient Education; Personalized e-Health; Natural Language Generation; Artificial Intelligence

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 important new approach in patient-centric care is through 'Information Therapy': prescribing the right information to the right person at the right time. Ideally, Information Therapy promises to deliver specific medical information to individual patients at just the right time to assist with decision-making or behaviour changes [5]. So far, however, the kinds of information delivered through specific patient portals or other individualized 'information prescriptions' is still quite generic (Claudette DeLenardo, Program Director, My CARE Source Patient Portal, Grand River Hospital, personal communication).

An effective means of providing patient-centric healthcare 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. As a key element in e-health services, personalized health education has great potential to provide relevant, patient-specific information that would integrate the clinical process with 'anywhere, anytime' healthcare delivery to produce both better-informed patients and enhanced patient outcomes. Some improvements in patient outcomes we may hope to see: greater compliance and satisfaction with therapeutic regimens, fewer complications, shorter hospital stays, and fewer return visits.

Value of Tailored Health Information

We are focusing 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, multistep 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. 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 multistep procedure, a handful of brochures would be required, which would lack cohesiveness and would likely be very confusing. Moreover, the complexity of the surgical procedures 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.

Our approach to personalization of health information is supported by studies in health communication which have shown that health-education material can be much more effective if it is customized for the individual patient in accordance with their medical conditions, demographic variables, personality profile, or other relevant factors. A recent special issue of *Patient Education and Counseling* (February 1999) reviewed key work in computer-generated tailoring of health information, including: smoking cessation (V.J. Strecher; W.F. Verlicer and J.O. Prochaska), nutrition (J. Brug et al.), and potential benefits of tailoring (D.S. Bental, A. Cawsey, and R. Jones; A. Dijkstra and H. De Vries). More recently, conferences in Artificial Intelligence and Computational Linguistics (*Workshop on Personalisation in e-Health*, 2006; *Argumentation for Consumers of Healthcare*, *American Association for Artificial Intelligence (AAAI) Spring Symposium*, 2006) have focused on personalization in e-health, attesting to its growing importance.

Natural Language Generation Systems for Personalized Health Information

The fundamental problem in personalization of patient education is that the process involves much more than just producing each brochure or leaflet in half a dozen different versions for different audiences. Rather, the number of different combinations of factors can easily be in the tens or hundreds of thousands. It is impossible to produce and distribute, in advance of need, the large number of different editions of each publication that is entailed by individual tailoring of health information.

Recently, researchers in Natural Language Generation have begun to apply methods from Artificial Intelligence and Computational Linguistics to develop automated systems for tailoring health information to individual patients (e.g., [1][4][6]). Our approach is similar in its underlying goal of providing more relevant, patient-centric health information through content geared to the individual. However, we aim to produce high-quality, finely tailored texts by avoiding the currently intractable problems inherent in generation 'from scratch'. We rely instead on the *pre-authoring* of a 'library' of reusable text, annotated with linguistic and formatting information, and a subsequent process of customization in which reader-appropriate pieces of text are selected, assembled, and automatically revised by computational 'text-repair' experts.

System Development Framework

The key components in our patient information tailoring system are:

- Creation of corpora of content variants written in language that will engage the patient and address individual concerns.
- A *Physician's Authoring Tool* that assists the physician in mapping from the various options at each stage of a medical intervention to corresponding content variations.
- A *Natural Language Generation Tailoring Engine* that will automatically select, assemble, and revise content from a library of reusable text to produce a customized version for an individual patient.
- A *Web-based framework* for delivery of tailored health education, e.g., a 'personal patient portal'.

1. Content Creation

Although patient educators now support the notion that personalization can enhance the uptake of information, bring about increased receptivity to behavioural change, and involve patients in their own healthcare decision-making, it is still the case that very little is as yet known about exactly *how* to customize language to gain these potential benefits. Even systems that purport to provide targetted health information generally do not provide truly personalized content. In a typical case, a patient may have her own 'patient portal', which will record her medications, treatments, appointments, etc., but the actual content she receives will still be generic. In the worst case, the patient might click on a link to gain information about her diagnosed cancer, and will receive a pop-up PDF document consisting of a lengthy brochure of 'boilerplate' text from a national cancer agency (Claudette DeLenardo, Program Director, My CARE Source Patient Portal, Grand River Hospital, personal communication).

We are working with a team of healthcare providers and medical language specialists to determine the specific factors that influence patient receptivity to health information and how these factors map to the actual language that should be used. This knowledge is being applied to create tailored corpora of patient-education content in various domains.

2. Physician's Authoring Tool

The classic bottleneck in the development of knowledge-based systems is *knowledge acquisition*, getting the knowledge from the domain expert, who is usually not a computer expert, into computational form. In the patient education case, authoring of materials has typically been accomplished through the interaction of the health professional with a 'knowledge engineer', someone trained in structured knowledge acquisition. But this makes the initial capture and on-going maintenance of the content dependent on a computer programmer, who may know little or nothing about medicine. Our goal is to provide a supportive authoring environment that will allow the domain expert—physicians and other health-care providers—to interact directly with the Tailoring Engine to enter the content variants that will later be used to produce personalized versions for individual patients.

3. The Tailoring Engine

In the past, researchers in Natural Language Generation (NLG) were generally limited to two main types of systems: simplistic approaches based on the use of canned text and syntactic templates, or more-complex systems that constructed text 'from scratch' using basic linguistic resources such as lexicons, grammars, and knowledge bases. In either case, however, the outcome was not ideal: either very simplified text or more-expressive output but with grammatical and stylistic faults.

NLG researchers are addressing this problem by *authoring* input for generation systems to produce expressive text without the accompanying faults of knowledge-based systems. Our solution has the best of both approaches, combining pre-authoring of input with sophisticated NLG techniques that can automatically edit a text to further refine it. We have developed a novel paradigm based on the concept of 'generation-by-selection-and-repair' [2], i.e., generation of new documents from pre-existing text through a process of reuse and revision. We start from an existing 'Master Document' that contains all the pieces of text that might be needed to tailor the document for any audience. Selections from the Master Document are made according to a individual reader profile, and then are automatically post-edited — 'repaired' — for form, style, and coherence.

4. Web-based Delivery of Health Information

Prof. Don Cowan and the Computer Systems Group (CSG) at the University of Waterloo have been developing portal technologies and deploying Web-based portals based on these technologies since the early 1990s. Current portals, called Community Learning Spaces, are constructed using the Web-based Informatics Development Environment (WIDE). The Community Learning Spaces are based on a service-oriented architecture and use the WIDE software toolkit to implement these services to enable the design, construction, deployment, maintenance and operation of complex Web-based systems. In the WIDE approach, "programming" has effectively been replaced with a XML-based declarative methodology, making it possible to provide a 'wizard' or forms-based approach to building Web-based systems. We are applying WIDE's software tools, services, and supporting frameworks to develop an integrated Web-based delivery mechanism for our Authoring Tool and Patient Information Tailoring Engine.

Project Development History

In 1994 the original "HealthDoc Project" [3] at the University of Waterloo was funded by Technology Ontario and the Information Technology Research Centre, the Province of Ontario's Centre of Excellence in Information Technology. The tailoring technology[1] developed in the original project was subsequently licensed in 1998 from our spin-off company, Inkpot Software Inc., to Health Media Inc. (Ann Arbor) and used in their commercial systems for both Web-based and print-based delivery of personalized health-education materials. In particular, our technology was used in a major commercial system that involved the automated generation of individually tailored Web pages on smoking cessation for patients taking an anti-smoking medication.

Project Team Roles and Competencies

The HealthDoc Project resumed in 2004, involving team members from the University of Waterloo, the Waterloo Institute for Health Informatics Research, the University Health Network of the University of Toronto, and the University of Southern California. Our project team includes a wide variety of academic, clinical, and health IT professionals fulfilling the following roles:

Health IT Leader: A professional who provides strategic leadership regarding all aspects of IT/IM for health organizations.

Healthcare Providers: Clinicians who provide specifications for design of our technology based on their day-to-day practices and who evaluate our systems and software tools.

Patient Education Providers: Experts in patient education are available to our project as advisors and evaluators through the University Health Network (University of Toronto) and Grand River Hospital (Kitchener-Waterloo).

Health Portal Developers: Experts in design and development of health portal systems and software tools.

Computational Linguists: Researchers in theory and design of Natural Language Generation systems.

Linguists: Researchers in rhetoric of medical language, physician/patient communication, and communication design.

Software Manager: Professional software developer who oversees all aspects of system design and development.

Technology Transfer: Advisors in academic/industry collaboration who provide liaisons with industry partners.

A Prototype Patient Information Tailoring System

The initial stage of our project focused on the creation of a corpus of patient-specific educational content by our surgeon team members covering a selected subprocedure of reconstructive breast surgery.

We developed an initial Web-based authoring environment that allows the medical domain expert to interact directly with the system from any provider location. The Authoring Tool allows clinicians to enter a spreadsheet of content, convert it to a standard Master Document, and view its discourse structure. Authors are thus given a means of displaying the complicated structure of the content in a form that makes subsequent editing much easier to understand. The updated version of the Authoring Tool allows direct creation of content in standard Master-Document format, then subsequent editing of both content variants and selection conditions for choosing particular variants.

To demonstrate and evaluate the results of our research so far, we implemented a pilot Web-based delivery system for generating tailored patient-education materials. The pilot system integrates the main components of our project—the corpus of content variants, the Authoring Tool, and the Natural Language Generation Tailoring Engine—with the Web Informatics Development Environment (WIDE). Several sample Master Documents for tailored health information (e.g., smoking cessation, reconstructive surgery) have been tested with the pilot tailoring system.

[1] A U.S. patent was granted on the HealthDoc 'Master Document' technology in March 2005, and the Canadian patent is pending.

Research Challenges

The next stage of our project is aimed at evolving each component of the current system:

Content creation. Reconstructive surgery will continue to be a very complex 'training' domain for the creation of personalized patient-education content. We will continue to investigate the rhetoric of physician/patient communication, including the following aims: to recognize various patient types, identify patient features which influence personalization of information, and investigate how these features relate to variations in both content and presentation (i.e., 'visual rhetoric') that will better engage patients and address their concerns.

Physician's Authoring Environment. The current generic authoring tool will be extended and enhanced to incorporate features geared to physicians: minimal input required from provider; use of medical 'metaphors' in interface design (e.g., physicians should be able to specify selection conditions on content variants in familiar medical manner of describing patient diagnosis); both 'push' and 'pull' methods of content authoring to allow author to directly enter procedure-specific text components (push) or use wizard-like query engine to intelligently prompt for details (pull); Web accessibility to permit use of the application independent of provider location. We will incorporate some handling of visual rhetoric and discourse plans into the authoring tool. We also plan to investigate the incorporation of higher-level physician and patient concerns (the 'knowledge level' described in [7]) and the modeling of workflow into the authoring environment.

Automated text revision. A major focus will be the investigation of architectures for Natural Language Generation systems based on 'text-to-text' generation, i.e., creation of new documents by reuse and revision of pre-existing texts. This will involve further development and subsequent integration of several components of our tailoring system: a 'deep' authoring tool for translation of surface text to linguistic representations used in complex repairs; repair visualization interface; individual text-repair engines; and planning mechanism for negotiating conflicts among repair engines.

"Write once, deliver anywhere". Our ultimate goal is the development of Web-based systems for generating personalized health information in which all aspects (authoring, generation of information) can be carried out over the Web and displayed on a variety of access devices (e.g., Web browsers, BlackBerry).

Acknowledgements
This work is supported by Bell University Laboratories and the Natural Sciences and Engineering Research Council of Canada.

The HealthDoc Project Working group:
S. Banks, P. Bray, H.D. Covvey, D. Cowan.
V. DiCiccio, C. DiMarco, O. Gladkov, R. Harris,
E. Hovy, J. Lipa. D. Mulholland, C. Schryer,
M. Skala, C. Strommer, J. Wang, J. Whissell
Address for correspondence:
cdimarco@uwaterloo.ca

References

[1] Bental DS, Cawsey AJ, and Jones R. Patient information systems that tailor to the individual. *Patient Ed. and Counselling*, 36, 1999.

[2] DiMarco C, Hirst G, and Hovy E. Generation by selection and repair as a method for adapting text for the individual reader. Flexible Hypertext Wkshop, ACM Hypertext Conf, 1997.

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

[4] Green N. Generation of biomedical arguments for lay readers. Intl Nat Lang Gen Conf, 2006.

[5] Mettler M and Kemper DW. Information Therapy: Health education one person at a time. *Health Prom. Prac.* 4(3), 214-217, 2003.

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

[7] Yang C. Design of a knowledge acquisition tool using a constructivist approach for creating tailorable patient education materials. MMath thesis, University of Waterloo, 2005.