

Applying document routing mode of information access in nursing diagnosis process

문서 라우팅 기법을 이용한 간호진단 과정에서의 정보접근

Woojin Paik, Konkuk University, wjpaik@kku.ac.kr

백우진, 건국대학교

Nursing diagnosis process is described as nurses assessing the patients' conditions by applying reasoning and looking for patterns, which fit the defining characteristics of one or more diagnoses. This process is similar to using a typical document retrieval system if we consider the patients' conditions as queries, nursing diagnoses as documents, and the defining characteristics as index terms of the documents. However, there is a small fixed number of nursing diagnoses and infinite number of patients' conditions in a typical hospital setting. This state is more suitable to applying document routing mode of information access, which is defined as a number of archived profiles, compared to individual documents. In this paper, we describe a ROUting-based Nursing Diagnosis (ROUND) system and its Natural Language Processing-based query processing component, which converts the defining characteristics of nursing diagnoses into query representations.

1. Introduction

According to North American Nursing Diagnosis Association (NANDA), a nursing diagnosis is defined as a critical judgment about individual, family, or community responses to actual or potential health problems or life processes. The goal of a nursing diagnosis is to identify health problems of the patient and his/her family. A diagnosis provides a direction for the following nursing care (Sparks et al, 2000). We are developing a Nursing

Diagnosis system, which is based on the document routing principles, to assist nurses in performing diagnostic tasks. The document routing is defined as automatic routing of incoming text documents based on content analysis of text (Guillermo, et al, 2000). The nurses are expected to enter the patients' conditions description composed on the basis of his/her observations and examinations and then receive a list of potentially relevant nursing diagnoses.

The nurses will verify each diagnosis and then select zero or more diagnoses as the final outcome.

2. ROUND System Overview

ROUND system is designed to process patients' conditions, which are entered on a computer terminal by nurses, as the incoming documents and the defining characteristics of the nursing diagnoses as the queries in a document routing system. The patients' conditions are matched against the query requirements, which are extracted beforehand from the defining characteristics. The matching process does not produce a ranked list of conditions. Instead, the matching process is defined as applying cutoffs to separate relevant descriptions from non-relevant descriptions (Croft, 1995). This means that the system will return a list of diagnoses, which are determined to be relevant to the patients' conditions. ROUND provides rich, deep processing of text by representing and matching documents and queries at the lexical, syntactic, and semantic levels, not simply by detecting the co-occurrence of words or phrases. Nurses are able to enter patients' descriptions as fully-formed sentences, with no requirement for special coding, annotation or the use of logical operators.

ROUND system is modular and performs staged processing of defining characteristics, with each module adding a

meaningful annotation to the text. For matching, a patient's condition description undergoes analogous processing to determine the requirements for matching.

3. Processing Patient Conditions

A preprocessor transforms raw digital data files of text into a uniform format suitable for further processing by the ROUND system. All texts are annotated with XML tags. The preprocessor identifies clauses and punctuation in a text, and annotates a document with identifying tags for these units. Conditions descriptions are first processed using an end-of-sentence detection program, followed by a part-of-speech (POS) tagger. In addition, hyphenated words are often given multiple tags—each constituent word is given a tag, and the whole hyphenated phrase is given a tag. Proper nouns and group common nouns (e.g., anti-cancer drugs) are recognized as important sources of information for detecting relevant documents in information retrieval (Liddy, et al, 1993).

Proper noun categorizer first locates the boundaries of proper noun phrases using the POS tags. Heuristics developed through corpus analysis are applied to bracket proper noun phrases, which contain embedded conjunctions and prepositions (e.g., Centers for Disease Control and Prevention). Proper noun categorization is the process whereby a proper noun is assigned to a single

category. Categories include drug name, treatment name, etc. To categorize a proper noun, proper noun suffixes, prefixes and infixes are examined for possible categorization information. Then, the proper noun is passed to a database to determine if an alternative, standard form exists. If the proper noun is an alias, the standard form is used for categorization. The proper noun is next run through context heuristic tests for possible categorization. Text-based clues are used for categorization. Proper nouns are compared to a database of significant personal first names for a possible match. Those proper nouns that remain uncategorized are assigned to the misc. category (Paik, et al, 1996).

Complex nominals (e.g., central nervous system) are important information-bearing phrases detected by the ROUND system and used in the conditions description-defining characteristics matching process. Complex nominal phrases are recognizable as adjacent noun pairs or sequences of non-predicating and predicating adjective(s) and noun(s). These pairs or sequences can be recognized from the output of the POS-tagged text in conjunction with various processing tools developed from corpus analysis.

The detection of complex nominals and proper nouns alone would not account for all of the information-rich content of typical English-language texts. Some

nouns, conflated nouns (e.g., inkwell), verbs, adverbs and adjectives also contain important information about the subject-contents of documents, and are detected by the single term detector. Numbers and numerically-related information (e.g., "\$" and other currency symbols) are also recognized.

4. Processing Nursing Diagnosis Defining Characteristics

After the defining characteristics is processed with the same Natural Language Processing modules as the patients' conditions descriptions, the characteristics are further processed with the sublanguage processor. Sublanguage processing is the beginning of a transition from a natural language query representation to a pseudo-logical representation of the query contents. The initial sublanguage processing of the query involves tokenization and standardization. After this initial processing, the defining characteristics are decomposed into a set of logical or pseudo-logical assertions by linking different portions of the query (i.e. various terms or groups of terms) with the operators. Thus, a series of operators are used to make these logical and pseudo-logical assertions. These operators relate terms and parts of the query text together, and also assign scores according to the formulas as

described below. Different operators assign different scores. 'Boolean AND' sums scores from ANDed terms and 'Boolean OR' selects the maximum score from all ORed terms

First, the Defining Characteristics Processor automatically constructs a logical representation of the natural language query. The user is not required to annotate the query in any way. A tree structure with terms connected by logical operators is constructed. An example of the result from processing the defining characteristics is shown below. It is one of 14 characteristics of 'Constipation' Diagnosis, which is related to gastrointestinal obstruction.

The Boolean logic representation of '*Abnormal tenderness or pain and feeling of rectal fullness or pressure*' is (*AND (OR 'abnormal tenderness', 'pain', 'feeling', (OR 'rectal fullness', 'rectal pressure'))*).

Various linguistic clues such as lexical clues and punctuation are used to determine the logical form of the query. The basis of this system is a sublanguage grammar which is based on the generalizations regarding the regularities exhibited in a large corpus of queries.

The sublanguage relies on items such as function words (articles, auxiliaries, and prepositions), meta-text phrases, and punctuation (or the combination of these elements) to recognize and extract the formal logical combination of relevancy

requirements from the query. In the example stated above, the positions and relations of the preposition "of", the conjunctions "and" and "or", and the comma and the period are used together to produce the appropriate logical relationship between the various items. The sublanguage interprets the query into pattern-action rules which reveal the combination of relations that organize queries, and which allow the creation from each sentence of a first-order logic assertion, reflecting the Boolean and other logical assertions or relations in the text.

5. Implementation and Evaluation

The development of the ROUND system is an ongoing effort. We developed a prototype system by converting 2,088 characteristics, which belong to 158 NANDA approved diagnoses, into logical requirements. In the first conversion attempt, the query processing sub-system of the ROUND system achieved 83% correctness rate. The automatic conversion results were compared to the manual conversion outcomes by considering the manual conversion is correct. The main cause of the errors was the complex nominal detection problems. Many boundaries of the complex nominals were not correctly identified due to the part-of-speech tagging problem. The second conversion

after adjusting the part-of-speech tagger achieved 94% correctness rate.

The patient conditions, which were used for the evaluation, came from 52 nursing care plans. The generation of the nursing care plans was a part of a course assignment for the 'Psychiatric Nursing' course, which was offered in the Spring and Fall 2005 semesters at the Department of Nursing Science, Konkuk University in Chungju, Korea. The course was for Juniors who were majoring in the nursing science. Each student developed a detailed case study report of one patient while the student was working as a student intern at a psychiatric ward for two weeks. The nursing care plan was one section of the case report. All case reports were mainly written in Korean with English translations for a number of important concepts. We manually translated the patient condition part of the nursing care plans to English as the ROUND system is designed to process English texts. We assumed that the instructor checked nursing diagnosis in the care plans as the correct diagnosis.

The precision of the ROUND system varies based on the cut-off criteria. By making the criteria high, the system achieves high precision but low recall. Inversely, the system achieves low precision and high recall by making the cut-off criteria lower. We are currently experimenting with the system to find the optimum cut-off criteria.

6. Conclusion

We have implemented an automatic nursing diagnosis selection system based on a document routing model. The nursing diagnosis defining characteristics are converted into extended Boolean standing queries by applying natural language processing techniques (Kendall, 1997). The patient conditions were treated as the incoming documents in a document routing system. The patient conditions were also processed by the natural language processing modules. The use of the extended Boolean queries allowed us to generate similarity scores between the nursing diagnosis defining characteristics and the patient conditions. In turn, we were able to use the cut-off criteria based on the similarity scores.

7. References

- Croft, B. 1995. What Do People Want from Information Retrieval?, *D-Lib Magazine*, 1. [online] <<http://www.dlib.org/lib/november95/11croft.htm>>.
- Guillermo R., Maria C.A., Pilar A., Adelaida G., Maria P.S., Susana S. 2000. An Architecture for Document Routing in Spanish: Two Language Components, Pre-processor and Parser. Proceedings of Second International Conference on Language Resources & Evaluation.

[o n l i n e]

<<http://citeseer.ist.psu.edu/348635.html>>.

Kendall, W.S. 1997. On some weighted Boolean models. In Jeulin, D. (Ed.) editor, *Advances in Theory and Applications of Random Sets*, Singapore. World Scientific Publishing Company.

Liddy, E.D., Paik, W., Yu, E.S. & McVearry, K. 1993. An overview of DR-LINK and its approach to document filtering. *Proceedings of the ARPA Workshop on Human Language Technology*, 358-362.

Paik, W., Liddy, E.D., Yu, E.S. & McKenna, M. 1996. Categorizing and standardizing proper nouns for efficient information retrieval. In Boguraev, B. (Ed.) *Corpus Processing for Lexicon Acquisition*. Cambridge, MA. MIT Press.

Sparks, S.M. and Taylor, C.M. 2000. *Nursing Diagnosis Reference Manual* 5th Ed. Pennsylvania: Springhouse.