Middle East Technical University

Department of Computer Engineering

- TURKUAZ -

Project Proposal

Tiran Software

Tahir Bilal
Onur Deniz
Soner Kara
M.Mert Karadağlı

Project Topic

Text-mining on Turkish radiology reports using Natural Language Processing and Ontology.

Project Description

The main purpose of this project is to extract meaningful data out of free-text radiology reports, so that the collected data can be easily manipulated and searched on demand. Text-mining on Turkish radiology reports is a challenging subject since hitherto there is not much research about text-mining on Turkish texts. In addition we will face with the complex structure of Turkish as well as hundreds of medical terms. On the other hand, the project will be very handful for academic use and it is an important research on automated medical information systems.

In order to extract high quality data out of free-text reports we have to choose the right text-mining techniques, some of which are natural language processing (NLP), machine learning and neural networks.

Initial Ideas

Until now we have done some research about text mining, radiology reports, NLP, machine learning, rule-based AI and other methods. Primarily we are planning to use NLP with medical ontology, and maybe some machine learning methods. With machine learning, in time the program may learn the relations of mostly used words in radiology reports and extract meaningful data. We plan to use Zemberek, an open source NLP library that can be very helpful for lexical analysis of Turkish reports. To identify medical terms, we are going to use SNOMED's semantic structure.

Our project will basically consist of following parts: A lexical analyzer, an NLP tool to analyze the sentences in the text syntactically, a procedure to reach SNOMED's terms and search word groups in the semantic structure, a procedure to analyze the texts semantically or to realize the semantic relations of the words in the text, a database to collect all the extracted useful data and a graphical user-interface.

Graphical user interface will be used to input the text documents to analyze. Additionally it will give the users the option to query the database that possesses the extracted useful data. These two tasks seem uncomplicated to implement, so we do not plan to arrange a project member only to program the GUI.

The database design part also doesn't seem to be very complicated, but significant indeed. A good database design is crucial not only for the efficiency of the database queries, and but also for the context and classification of the extracted data.

None of us took a course on NLP or artificial intelligence before, so we still need to do a lot of research about these topics altogether, thus we do not plan to distribute the main tasks so early among group members. In first place, all the group members must attain a rather profound apprehension of NLP, text mining and artificial intelligence.

Ground Rules

1. Members have to attend to weekly meetings.
2. The penalty for not attending a meeting is ordering meals at Çarşı for other group members.
3. Meetings can be postponed if the meeting day runs across a very serious exam, and all other members agree. The delay may be utmost 3 days.
4. On legal holidays, if members are not in Ankara they may attend the meeting via web.
5. Members have to accomplish jobs which will be assigned during meetings.
6. All members have to check their e-mail inboxes at least twice a day.
7. If a member has been asked a question on the yahoo group, he must answer it in two days.
8. New ground rules may be added or the existing rules may be changed by consensus.

Group Members

Name	Bilal, Tahir
ID	1394741
e-mail	tahirbilall@gmail.com
Phone	+90 555 309 9648
Roles	Devil's advocate, gatekeeper
Expected grade	BA

Name	Deniz, Onur
ID	1391002
e-mail	denizonur@yahoo.com
Phone	+90 535 855 4320
Roles	Leader, recorder
Expected grade	BA

Name	Kara, Soner
ID	1395110
e-mail	e139511@metu.edu.tr
Phone	+90 532 645 6467
Roles	Initiator, timekeeper
Expected grade	BA

Name	Karadağlı, Mustafa Mert
ID	1395128
e-mail	m_mertkaradagli@yahoo.com
Phone	+90 505 451 3447
Roles	Initiator, dreamer
Expected grade	BA

Communication Method

Since we are taking common courses, we are going to communicate face to face on daily basis. Our weekly meeting is planned to be on Tuesday between 15:30-16:30. If there is a need on Friday at 13:30 we have a extra meeting.

Most of the time, the members are online. So, MSN is another communication option for daily communication. We have a Yahoo Group (tiransoft) for long term communication and to share private links and files with members of the group.

We also have a web page to announce our progress publicly located at the address http://www.metu.edu.tr/~e139474/

Risks Involved

There are a couple of risks that are involved in this Project:
- Deadline may not be met. There will be penalties to overcome this.
- One or more group members may not continue academic classes for a while due to illness. This may need task redistribution
- There may be some conflicts on a subject among the members and a proper decision may not be made for a while. In this case a great deal of effort will be given to reach a solution.