MIDDLE EAST TECHNICAL UNIVERSITY

DEPARTMENT OF COMPUTER ENGINEERING

'Text Mining On Turkish Medical Radiology Reports'

Project PROPOSAL

By

Fall, 2007

Table of Contents

1. Team Information

1.1 Company Name

The name of our company is "SBAYazılım".

1.2 Company Members

SBAYazılım Team: sbayazilim@yahoo.com

SBAYazılım Website: http://www.ceng.metu.edu.tr/~e1395383/sbayazilim

Name: Kerem Hadımlı

E-mail: wastiee@gmail.com

Phone: 0536 352 59 94

Name: Çiğdem Okuyucu

E-mail: e1448976@ceng.metu.edu.tr

Phone: 0539 254 27 10

Name: Makbule Gülçin Özsoy

E-mail: e1395383@ceng.metu.edu.tr

Phone: 0535 647 39 37

Name: İpek Tatlı

E-mail: e1395557@ceng.metu.edu.tr

Phone: 0537 774 42 24

1.3 Member Roles

In our project team there will be;

- a project leader who is responsible for the general coordination of the team,
- a recorder who is responsible for recording the points discussed in the meetings,
- a timekeeper who keeps the track of time spent on the meeting,
- an initiator who initiates the meeting and suggest new ideas,
- a gatekeeper who provides that each member states his/her ideas,
- a contact person who is responsible for the meeting with business contacts,
- a devil's advocate who tries to find out missing and faulty points,
- a summarizer who summarizes the ideas discussed during the whole meeting.

These roles are distributed among the members as follows:

Kerem Hadımlı: Team Leader, Initiator
Çiğdem Okuyucu: Gatekeeper, Contact Person
Makbule Gülçin Özsoy: Devil's Advocate, Recorder
İpek Tatlı: Summarizer, Timekeeper

1.4. Meeting Schedule

We all have had others´ *cellular phone numbers* and *instant messaging addresses* for communication. We also have a *e-mail group* for sharing ideas, attachments and each meeting's decisions; and a *wiki* for sharing documents and sources for our research.

We have twenty – minute long *meetings with Umut Eroğul* every Thursday between 13.40 and 14.00. We also have *another two meetings* with team members every week. They are on Wednesdays between 16.00 and 18.00 (for discussing future work); and on Tuesdays between 18.00 and 20.00 (for review of last week). We have arranged them according to our weekly schedule and free time. We will have two – hour long meetings at minimum but they can be extended.

1.5. Our Expectations

We have thought about our expectations from this project during our meetings. Majority expects some grades between BB and BA, but we will work on this project much more than the grades than we expect. Then maybe we will deserve higher grades.

Kerem Hadımlı:	BA – AA
Çiğdem Okuyucu:	BB – BA
Makbule Gülçin Özsoy:	BB – BA
İpek Tatlı:	BB –BA

1.6. Risk Management

Risk	Probability	Impact	Risk Mitigation Monitoring and Management
Lack of required skills	20%	4	Increase the workloads of team members to acquire these skills in time.
Unexpected overloads	70%	4	Manage time well and increase work.
Poor time management	50%	4	Observe the improvement of the project every week trying to obey the schedule. Accordingly, work harder or contact the instructor and drop some features of the final deliverable product if acceptable by the instructor.
Temporarily unavailability of one or more of the team members	5%	3	Increase the workloads of other team members.
Severance of a team member	5%	3	Increase the workloads of other team members.
Lack of knowledge about application area	60%	4	Study application area thoroughly.
Polarization between team members	10%	4	Solve the problem anyway.
Physiological problems	40%	2	Show more tolerance to that member.
Irresponsibility of members	15%	4	Show no tolerance. Give some reasonable punishment.
Missing the meetings twice.	10%	2	Give some reasonable punishment.
Missing deadlines.	50%	4	Increase the workloads of team members.
Treating rude and being offended to other members	10%	4	Behave as nothing has happened during the team work & meetings.

2. Project Information

2.1 Project Title

The title of our project is "RadioRead".

2.2 Project Description

Medical imaging and reporting is getting more importance in healthcare services. Quality of imaging techniques and quality of reporting have big roles on the diagnostic success of radiological imaging. Interpretation of radiological images is done by doctors in an intuitive way for accurate analysis of images.

In current approach, radiology reports are stored as free-text in hospital archives. The process is usually done in the following order:

1. Radiologists examine patients' films and medical images and record their findings loudly to a tape

2. Secretaries listen to the tape and transfer them to text documents. These documents are stored in hospitals' archives to be later reviewed when necessary.

The problematic part in this approach is that it makes statistical analysis and retrieval of similar data unfeasible. To solve this problem, there is a need to get information from these free-texts by using text-mining approaches.

2.3 Project Aim

In this project we will try to get meaningful data from Turkish radiology reports and store these data in a proper way which will be fast and reliable to retrieve. Radiology reports are free-texts which have no formatting except the rules of Turkish.

Our team expects to provide a way to

- find similar patients, diseases,
- categorize diagnosis,
- identify integrity of treatments and variances between patients with the implementation of this project.

2.4 Initial Ideas and Technology

We are planning to use *Natural Language Processing (NLP)* and *Artificial Intelligence (AI)* techniques. We will use *Java* as implementation language, and will utilize *Zemberek Library*[1], *TDK online*[2] and *Snomed database*[3].

2.5 Software Development Method

We plan to use a mixture of Waterfall and Evolutionary development methods.

3. References

[1] Zemberek, http://code.google.com/p/zemberek/

[2] Türk Dil Kurumu Online Word Catalog, http://www.tdk.gov.tr

[3] Snomed Medical Terminology Database, http://www.snomed.org