

**Middle East Technical University**

**Department of Computer Engineering**



**-TURKUAZ PROJECT-**

**TIRAN SOFTWARE  
REQUIREMENT ANALYSIS REPORT**

**RadeX**

*Tahir Bilal  
Onur Deniz  
Soner Kara  
Mert Karadağı*

## TABLE OF CONTENTS

---

Table Of Contents .....	2
1. Project Overview.....	4
1.1.Problem & Project Definition .....	4
1.2.Project Title.....	4
1.3.Project Scope .....	5
1.3.1. Best Case .....	5
1.3.2. Worst Case.....	5
1.4.Process Model .....	5
2. Research .....	6
2.1. Literature Research .....	6
2.1.1. A Statistical Language Processor for Medical Reports.....	6
2.1.2. A General Natural Language Text-Processor for Clinical Radiology.....	6
2.1.3. Medical Diagnosis with C4.5 Rule Preceded by Artificial Neural Network Ensemble .....	7
2.2. Market Research.....	7
2.2.1. MedIE (MEDical Information Extraction).....	7
2.2.2. HITEX.....	7
2.3. Technical Research .....	8
2.3.1. Zemberek .....	8
2.3.2. DBMS Selection.....	8
2.3.3. Machine Learning Methods.....	8
2.3.3.1. Genetic Algorithms.....	8
2.3.3.2. Artificial Neural Networks.....	9
2.3.3.3. Decision Tree Learning .....	9
2.3.3.4. Ruled-based AI Methods.....	9
3. REQUIREMENTS .....	9
3.1. System Requirements .....	9
3.1.1. Software Requirements .....	9
3.1.2. Hardware Requirements.....	9

3.2. Functional Requirements .....	10
3.3. User Requirements.....	10
4. Modeling .....	11
4.1. Data Modeling.....	11
4.1.1. ER Diagrams.....	11
4.1.2. ER Diagrams Explanation .....	14
4.2. Functional Modeling .....	14
4.2.1. Data Flow Diagrams .....	14
4.2.2. Data Flow Explanation .....	17
4.2.2.1. Pre-processor .....	17
4.2.2.2. Semantic Association .....	17
4.2.2.3. Learn and Extract.....	18
4.3. Behavioural Modeling .....	18
4.3.1. State Transition Diagrams .....	18
5. Scheduling.....	20
6. Planning.....	20
6.1. Risk Plan .....	20
6.2. Quality Plan .....	20
6.3. Test Plan .....	20
6.4. Evaluation and Validation Method.....	21
7. Appendix.....	21
7.1. Data Dictionary .....	21
7.1.1. Database Tables .....	22
7.2. Gantt Chart.....	23

# 1. PROJECT OVERVIEW

---

## 1.1. PROBLEM & PROJECT DEFINITION

---

The invention of computers and the advancement in data storage technologies flourished the use of electronic documents to store data. Since electronic documents have so much advantages over manuscripts or typewritten documents, this was a profound technology revolution that had a huge impact in our lives.

Yet, electronic documents don't reveal the information they have immediately, a human being still has to read a free-text electronic document to comprehend its contents. This is a bottleneck for information identification and association of related information. To be able to maximally make use of the electronic platform, information should be easier to obtain, search and identify.

There is a great attempt in this trend in a number of ways. Semantic web technology can be given as an illustrating example. Semantic web tries to extend the web to such an extent where content can be expressed not only in natural language, but also in a format that can be read and used by software agents, thus permitting them to find, share and integrate information more easily.

As in semantic web, the clinical records and reports of patients contain much potentially useful information in free text form that is not directly searchable. By extracting useful data from clinical reports, records of patients can be held at databases, which will drastically help the diagnosis of further or future clinical problems of patients. Moreover new correlations about some illnesses or drugs such as a drug's unnoticed side affect could be more easily discovered. Such advancement will help the medical science and diagnosis a lot.

Our project topic is text-mining in a specific medical branch's texts, Turkish radiology reports. 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 there is not much research about text-mining on Turkish texts. Additionally 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 handfull 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 such as specific natural language processing and machine learning methods.

## 1.2. PROJECT TITLE

---

Our radiology report analyzer is named **RadeX**, an abbreviation for Radiology Data Extractor.

## 1.3. PROJECT SCOPE

---

RadeX will consist of two main components, first of which is radiology report analyzer/information extractor, and the second one is a database query interface.

- The radiology report analyzer/information extractor component is the core component of our system. Our project's whole scope depends on the scope of this component. This component will be used to analyze a radiology report and extract data from it. The extracted data will be viewed on screen; it will be possible for the user to edit the viewed information. Moreover, this component will make it possible to insert the extracted data to a database upon request.
- The database query interface will be used to query the database.

---

### 1.3.1. BEST CASE

---

In the optimum case, our report analyzer component will be able to analyze radiology reports with up to 95% accuracy. It will be able to associate semantics for nearly all medical terms. The Turkish terms won't cause a problem, because by means of an internal spell checker it will be able to deduce the corresponding Latin terms. Moreover, the problems posed by complex Turkish grammar will be overcome. There are different types of radiology reports. If another type of radiology document report shows up, it will be able to successfully learn the inner format of that report type upon analyzing lots of documents of that type.

---

### 1.3.2. WORST CASE

---

In the worst case, our report analyzer component will be able to analyze radiology reports with at least 75% accuracy. If our learning algorithm doesn't succeed as we expected and the Turkish medical terms used in the documents don't converge to a number that we can overcome, this may be the case.

## 1.4. PROCESS MODEL

---

We are planning to use the incremental model. We have a set of fixed requirements and a strict purpose, we need to extract information from free-text radiology reports and because of these facts using spiral model would be unnecessarily exigent. We do not plan to return to requirement phase all the time and check whether requirements are satisfied or not in order to continue subsequent phases. Nevertheless, the waterfall model is also not very appropriate for us, since we have a lot of design options to implement some parts of the project, for machine learning phase of our project, we may use a neural network method, or we may use decision trees, and maybe both. Another unclear point about our project is the use of SNOMED. In the radiology reports that we researched on, the medical terminology didn't seem to be written completely in Latin, rather most of the Latin words were somewhat adopted to Turkish. These jumble words are neither available in Turkish dictionaries nor in Latin nor English. Due to these puzzling in specifications using a sequential work without never looking back and revising also seems not possible. We will use the incremental model, and try to deliver a core-product first, with text analysis, user interface and database querying capabilities. These components do not seem to be changed in the future, yet the technologies we use for extracting meaningful data and for machine learning may be changed, so we will incrementally enhance RadeX.

---

## 2. RESEARCH

---

### 2.1. LITERATURE RESEARCH

---

We carried out a lot of literature survey. In this way, we have learnt current methods which are being used in text mining nowadays. Moreover, latest developments in this field can be observed easily by reading academic papers. Here we give some of the papers we read for our literature survey.

---

#### 2.1.1. A STATISTICAL LANGUAGE PROCESSOR FOR MEDICAL REPORTS

---

In this paper, authors present their experience about applying a statistical NLP system for radiology reports with focus on sentence parsing and semantic interpretation. The idea of the parser is based on dependency diagram via words. For example, when “The large mass” word phrase is interpreted with this method, an arc from word “the” to “mass” and also arc from word “large” to “mass” indicates that words “the” and “large” modifies word “mass”. Authors present a statistical NLP processor based on resonance probabilities between word pairs. Parser gathers word affinity knowledge from training sentences, whose dependency diagrams are manually specified. The performance of the parser is measured with 90% recall (percentage of correct arcs the parser identifies) at 89% precision (percentage of arcs reported that are correct).

This paper shows how statistical NLP can be applied to our project and guarantees a high (acceptable) performance can be obtained with dependency diagram method.

*A STATISTICAL LANGUAGE PROCESSOR FOR MEDICAL REPORTS*  
Ricky K. Taira, Phd and Stephen G. Soderland, Phd  
Proc AMIA Symp 1999:970 -974

---

#### 2.1.2. A GENERAL NATURAL LANGUAGE TEXT-PROCESSOR FOR CLINICAL RADIOLOGY

---

Authors developed and evaluated a text processor that extracts and structures clinical information from textual radiology reports and translates the information to terms in a controlled vocabulary so that the clinical information can be accessed by further automated procedures. Besides the papers methodology and general concepts about NLP mentioned in this paper, this paper contributes us with its semantic classifications used to differentiate medical terms. A few of these classifications are listed as examples:

**Bodyloc:** Terms denoting a well-defined area of the body or a body part (e.g. hilum, left lower lobe, carotid artery)

**Device:** Terms denoting surgical devices that are evident on the radiology report. (e.g.surgical wires, swan ganz catheter)

**Disease:** Terms denoting a disease. These terms are based on the disease axis in SNOMED^ (e.g. asthma, cardiomyopathy)

*A GENERAL NATURAL LANGUAGE TEXT-PROCESSOR FOR CLINICAL RADIOLOGY*  
Carol Friedman, PHD, Philip O. Alderson, MD, John H. M. Austin, MD, James J. Cimino, MD, Stephen B. Johnson, PHD  
Journal of the American Medical Informatics Association, 1, 2, (1994), 161--174  
1994 Mar-Apr

---

### 2.1.3. MEDICAL DIAGNOSIS WITH C4.5 RULE PRECEDED BY ARTIFICIAL NEURAL NETWORK ENSEMBLE

---

This paper demonstrates a combination of artificial neural networks and decision trees to make medical diagnosis. At first, an artificial neural network is trained to generate a data set. Then, a new training data set is generated by feeding the feature vectors of the original training instances to the trained data set. Next, the expected class labels of the original training instances are replaced with the class labels outputted from the trained ones. Finally, a specific rule induction approach, i.e. C4.5 Rule, is used to learn rules from the new training data set. Case studies on diabetes, hepatitis, and breast cancer show that this method could generate rules with strong generalization ability and strong comprehensibility.

This paper shows us that a technique composed of neural networks and C4.5 can be used to make medical diagnosis.

*MEDICAL DIAGNOSIS WITH C4.5 RULE PRECEDED BY ARTIFICIAL NEURAL NETWORK ENSEMBLE*  
Zhi-Hua Zhou; Yuan Jiang  
*Information Technology in Biomedicine, IEEE Transactions on*  
*Volume 7, Issue 1, Mar 2003 Page(s): 37 - 42*

---

## 2.2. MARKET RESEARCH

---

In order to gain some insight about what can be done about text mining in medical reports, we searched the web for similar programs. We couldn't find any fully functional Turkish NLP program in this field. However, there are several successful programs for English. We examined them to find out a general architecture for our project. Here we describe briefly some of them.

---

### 2.2.1. MEDIE (MEDICAL INFORMATION EXTRACTION)

---

This project which is conducted in Drexel University, aims to extract patient information with breast complaints from free-text clinical records. It is a part of medical text mining project. Three approaches are used to solve information extraction tasks:

- \*graph-based approach: It uses the parsing result of link-grammar parser. It was invented for relation extraction and high accuracy was achieved.

- \*ontology-based approach: It is simple and efficient and focuses on extracting medical terms of interest.

- \*NLP-based and ID3-based decision tree coupled: Performs text classifications.

---

### 2.2.2. HITEX

---

HITEX (Health Information Text Extraction) is an open-source NLP software application, which extracts key findings for a research study on airways disease and smoking status extraction. It was developed by a group of researchers at the Brigham and Women's Hospital and Harvard Medical School.

HITEX is built on top of Gate framework. It consists of 11 Gate modules, which are: 1- Section splitter, 2- Section filter, 3- Sentence Splitter, 4- POS Tagger, 5- Noun Phrase Finder, 6- UMLS concept mapper, 7- Negation finder, 8- N-gram tool, 9- Classifier and 10- Regular expression-based concept

finder, 11- Sentence Tokenizer.

All or any combination of these components can be used to solve a given problem, which enables the program to adapt to different tasks easily.

We couldn't try this program due to lack of UMLS. Nevertheless, we were inspired a lot from the structure, components and general flow of the program by reading documents provided.

## 2.3. TECHNICAL RESEARCH

---

In the course of this project we will deal a lot with Natural Language Processing. NLP is a subarea of artificial intelligence that deals with analysis, manipulation and generation of natural language texts.

We are planning to implement RadeX in Java platform. There are lots of reasons of this choice. First of all, all of us are comfortable with Java. Secondly Java has a huge library support. Since Java has a very extensible structure and since it is a high level programming language it is quite popular these days.

---

### 2.3.1. ZEMBEREK

---

Zemberek is an open source NLP tool for Turkish. It is a successful tool for morphological analysis and part of speech tagging. Zemberek is written in Java, and adequate documentation exists to make use of it as a Java library. Zemberek is another reason that we chose to program in Java.

---

### 2.3.2. DBMS SELECTION

---

We need to use a DBMS in our project. At this moment, we are planning to use MySQL since it is relational, free, and it has the most significant attributes of a DBMS. In the second semester, we will assign the task of comparing DBMSs and selecting the appropriate one for our project to one member of our group. After taking feedback from that member, we will consider changing the DBMS that we use.

---

### 2.3.3. MACHINE LEARNING METHODS

---

We will definitely use one or more machine learning techniques in this project to automate the learning process of unknown words. Machine learning is a broad subfield of artificial intelligence which is concerned with the design and development of algorithms and techniques that allow computers to "learn".

---

#### 2.3.3.1. GENETIC ALGORITHMS

---

Genetic algorithms are machine learning techniques to provide learning by an analogical model to biological evaluation. In order to find an optimal solution to a problem genetic algorithms generate successor possible solutions by repeatedly mutating or recombining parts of the best currently known solutions. GAs(genetic algorithms) have a terminology similar to biology comprising of crossover, mutation, population, hypotheses, chromosome.



#### 2.3.3.2. ARTIFICIAL NEURAL NETWORKS

---

Artificial neural networks are mathematical computational learning models based on biological neural networks. The study of artificial neural networks has been inspired by the observation that biological learning systems are built of very complex webs of interconnected neurons. The human brain is estimated to contain an interconnected network of approximately  $10^{11}$  neurons. Likewise their biological counterparts, artificial neural networks are comprised of interconnected units. Instead of neurons, these are mathematically modeled ones like perceptrons and sigmoids. These units take real-valued inputs and by means of a linear or non-linear computation produce a single real-valued output.

#### 2.3.3.3. DECISION TREE LEARNING

---

Another machine learning technique is decision tree learning. Decision tree learning is a method for approximating discrete-valued target functions in which the learned function is represented by a decision tree. Learned trees can also be viewed as sets of if-then rules. Decision tree learning terminology is not as complicated as genetic algorithms' or neural networks'. Another good point about decision tree learning methods is that they are mostly applied for text-base learning such as diagnosing medical cases from reports.

#### 2.3.3.4. RULED-BASED AI METHODS

---

Ruled-based AI methods are another option. In rule-based AI, a knowledge base holds a set of if-then-else rules that specify how to act on assertions which are about the current knowledge. 'working memory' holds a set of assertions. Then the rules in knowledge base are exercised in working memory to evaluate the assertions.

## 3. REQUIREMENTS

---

### 3.1. SYSTEM REQUIREMENTS

---

In order for our program to run gracefully, system should have these software and hardware requirements:

#### 3.1.1. SOFTWARE REQUIREMENTS

---

- MySQL Server
- Java Run Time Environment
- Windows Environment
- Internet Connection

#### 3.1.2. HARDWARE REQUIREMENTS

---

- 1.5 GHz CPU
- 512 MB Ram
- 1 GB Hard Disk

### 3.2. FUNCTIONAL REQUIREMENTS

---

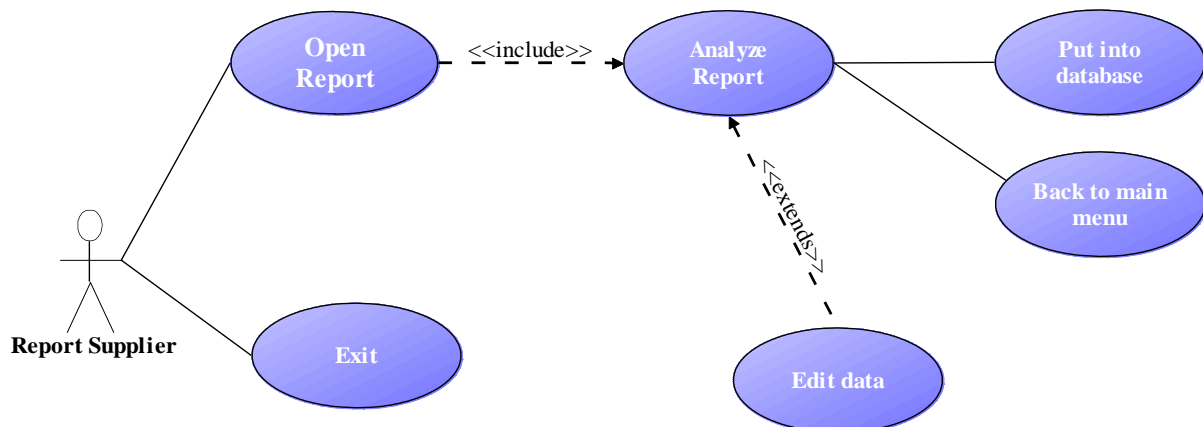
The program must satisfy the following functional requirements:

1. Program should extract meaningful data from reports. The decision about what information is to be extracted and is important should be made with learning techniques using statistical and ontological data acquired from all reports.
2. Program should store meaningful information in the database.
3. Program should integrate extracted information with existing information system to make all of the records searchable.

### 3.3. USER REQUIREMENTS

---

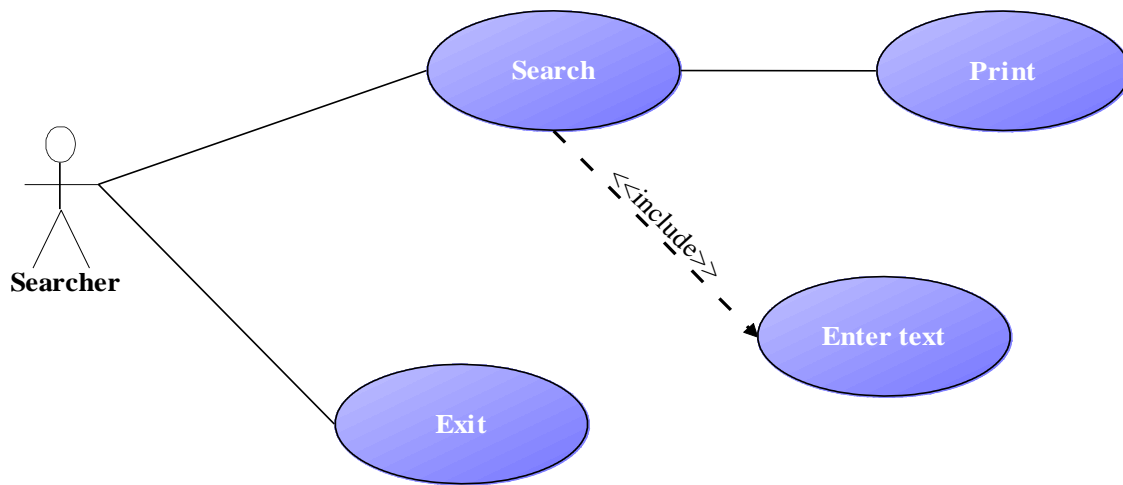
**USE CASE DIAGRAM: Report Supplier/Analyzer**



#### ***User 1 (Report supplier) scenario***

1. The user can choose "open report" from main menu.
  - 1.1. When it's opened, it's shown on the screen.
  - 1.2. Now the user can choose "analyze report" to extract data.
    - 1.2.1. When analyzing completed, extracted data are displayed to user.
    - 1.2.2. Extracted data can be edited manually.
      - 1.2.2.1. User can accept the changes.
      - 1.2.2.2. User can cancel the changes.
    - 1.2.3. User can send the extracted data to database.
    - 1.2.4. Back to menu
2. Exit

## USE CASE DIAGRAM: Searcher



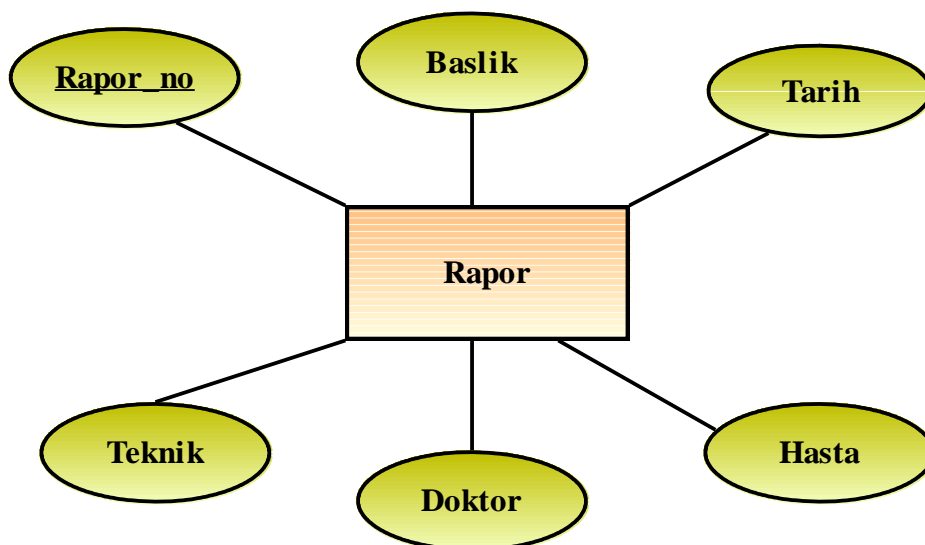
### User 2 (Searcher) scenario

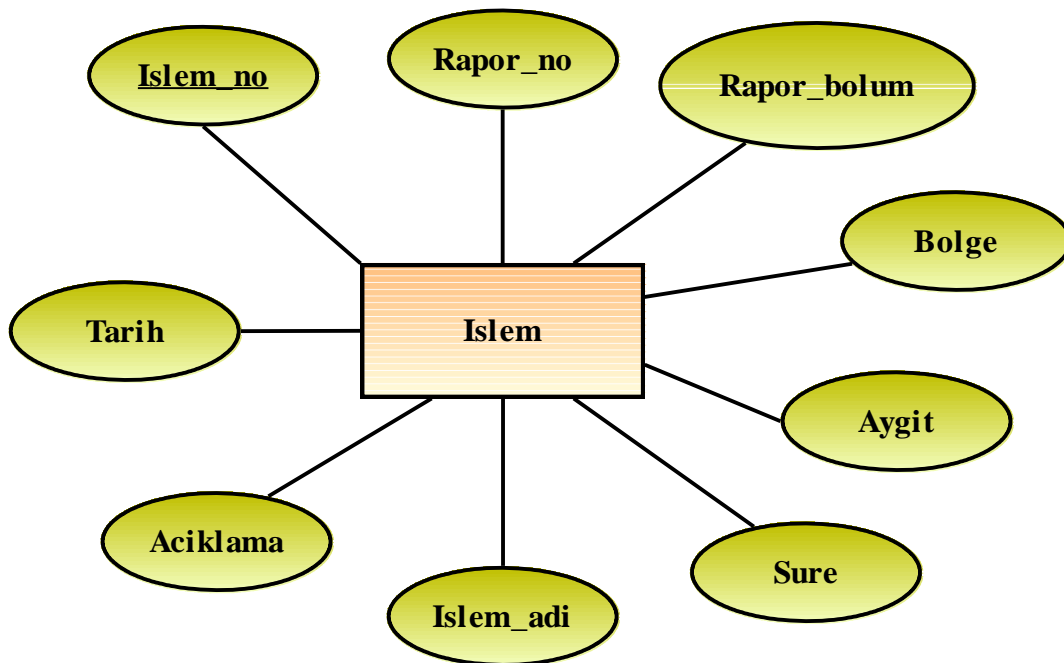
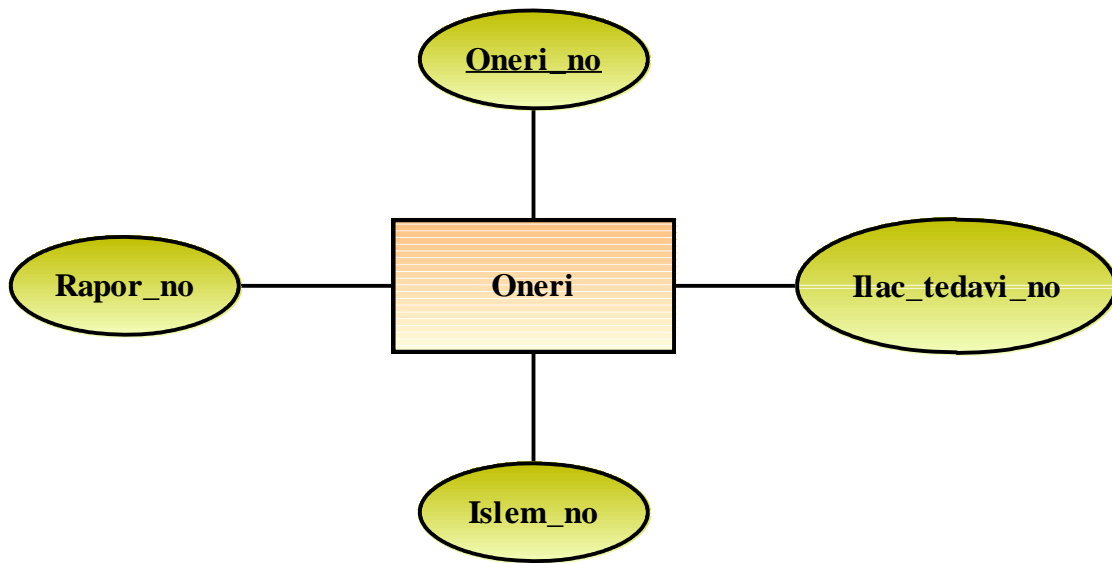
1. Search: - The user chooses “search” from main menu.
  - 1.1. Enter text: - A text box appears, in which query string will be written.
  - 1.2. When the search is over, results are shown on the screen.
  - 1.3. Now, the user has the option to print out the results.
2. Exit

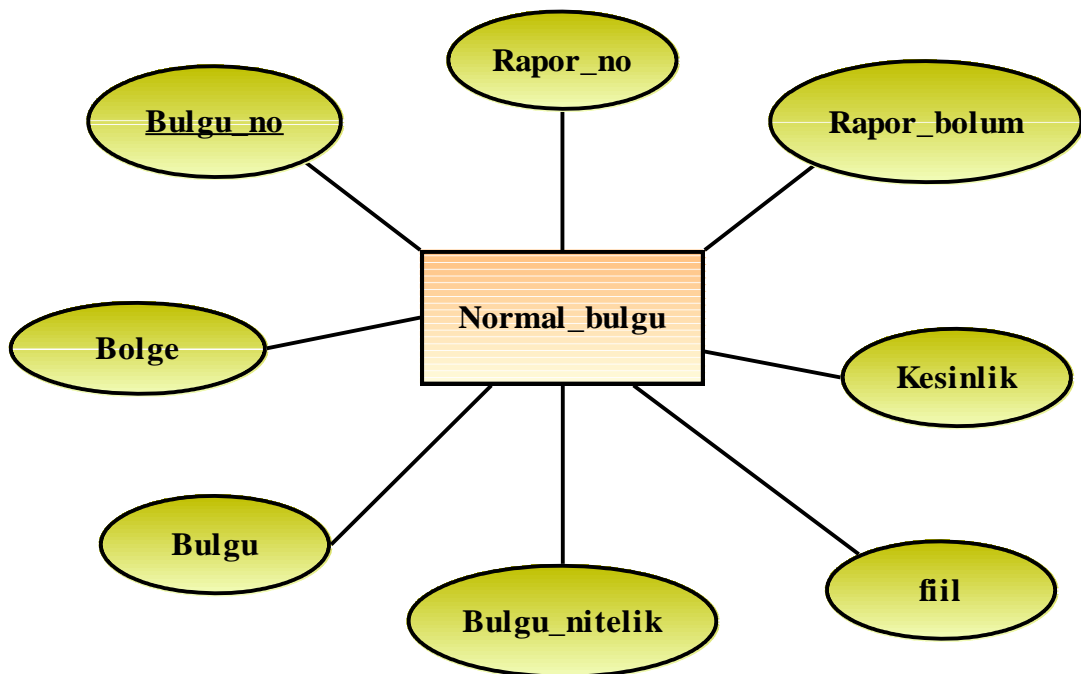
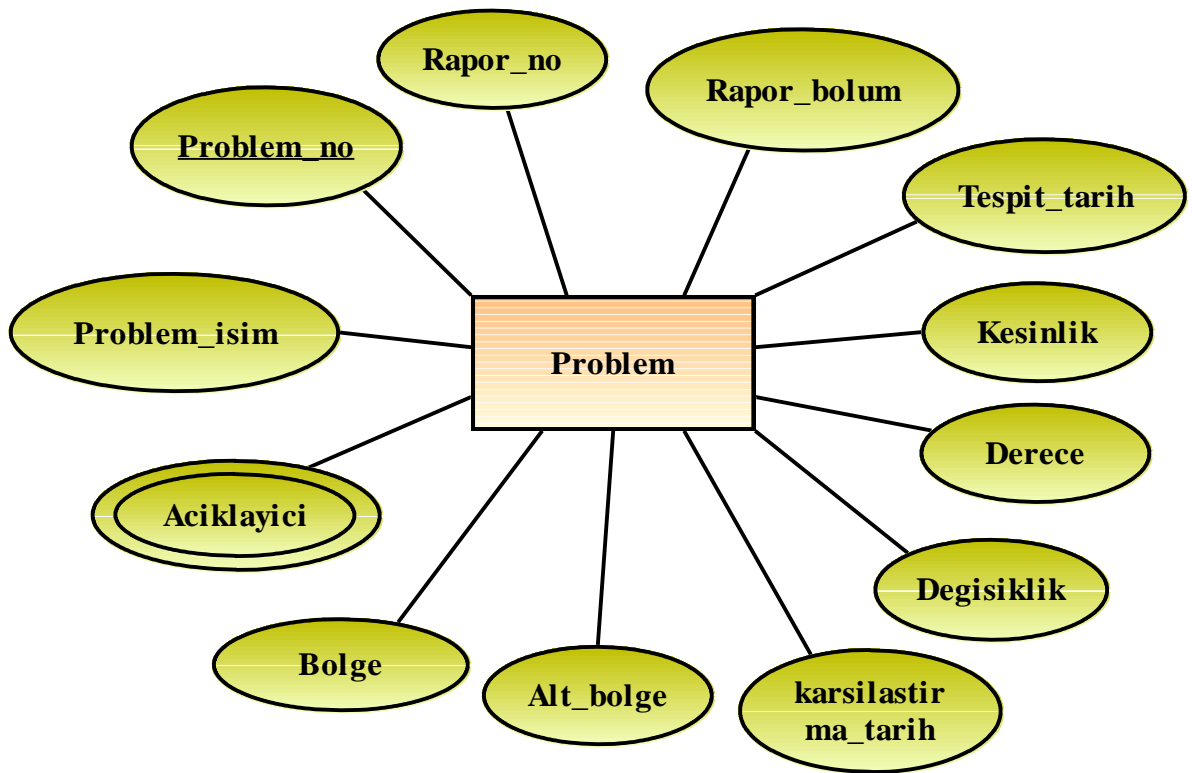
## 4. MODELING

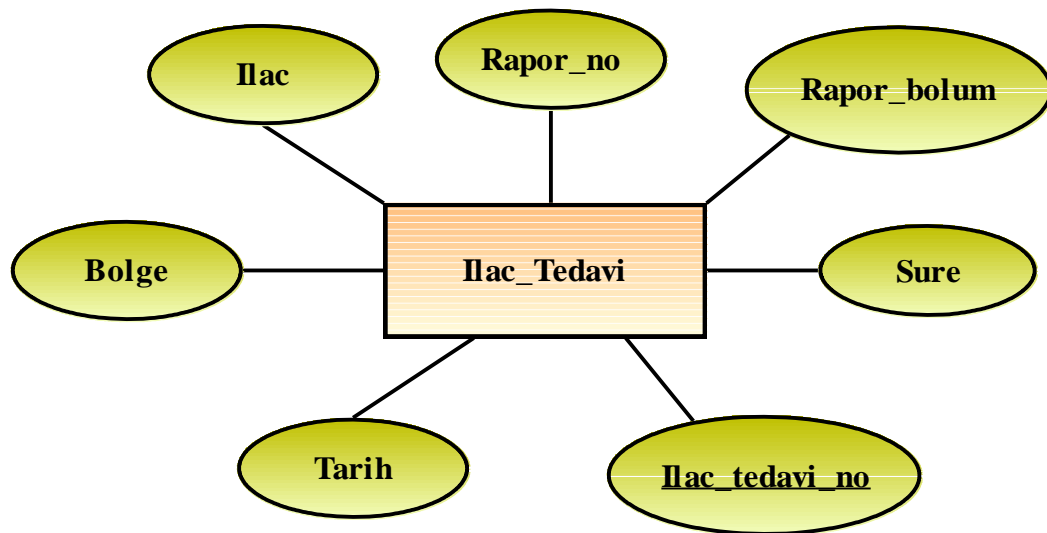
### 4.1. DATA MODELING

#### 4.1.1. ER DIAGRAMS







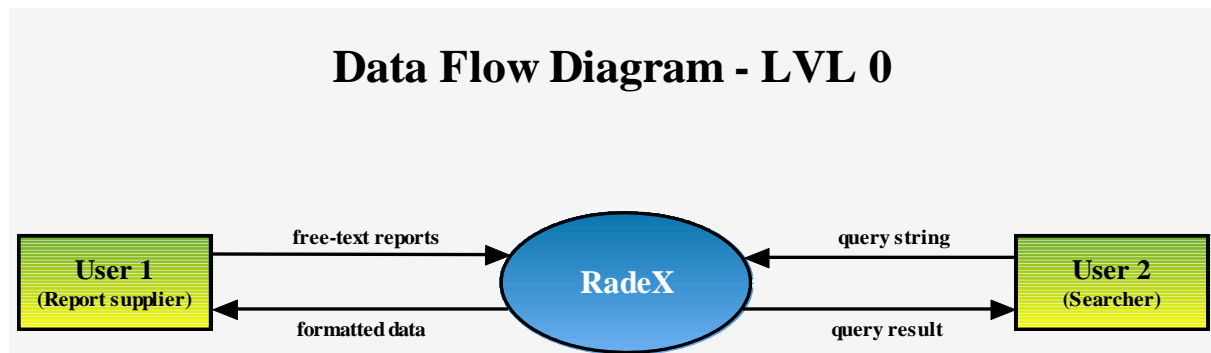


#### 4.1.2. ER DIAGRAMS EXPLANATION

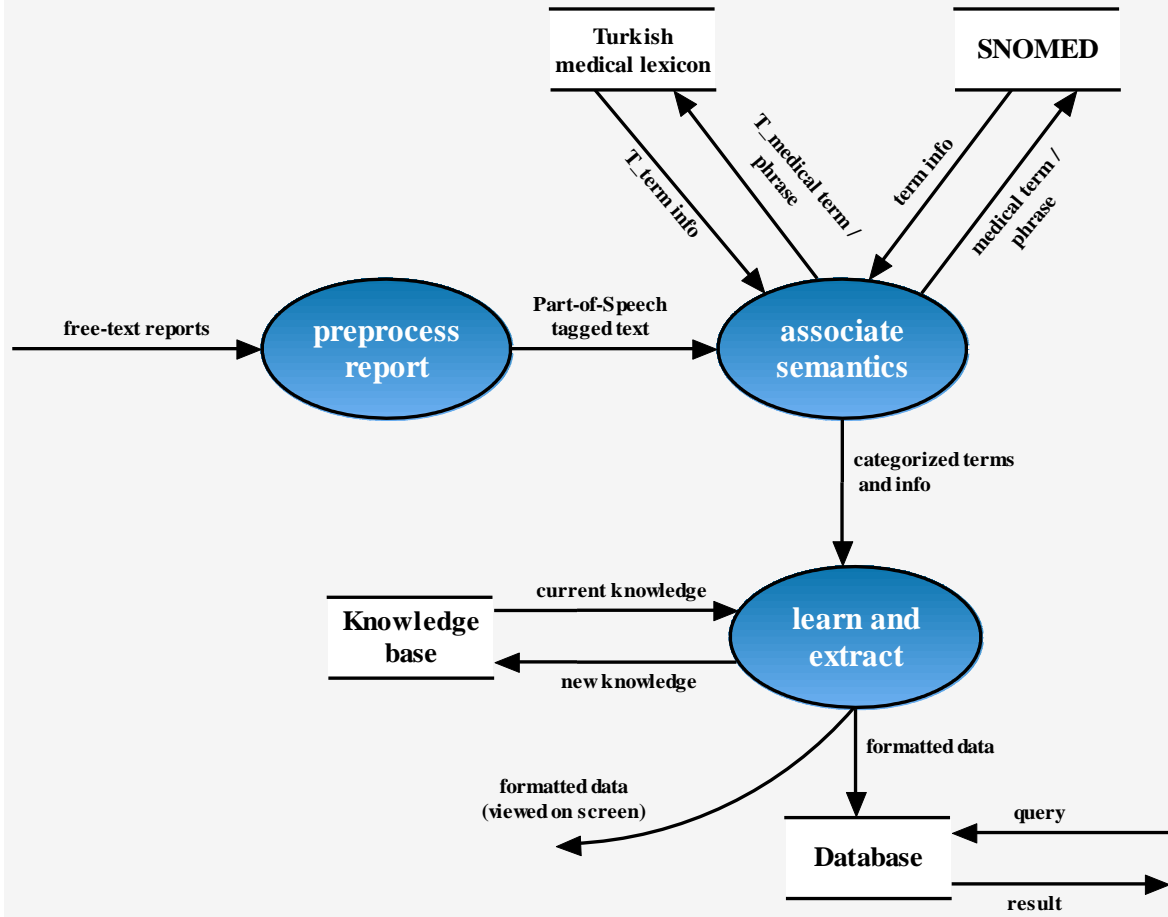
Look at the data dictionary at Appendix 7.1.1. for the explanation of database tables.

## 4.2. FUNCTIONAL MODELING

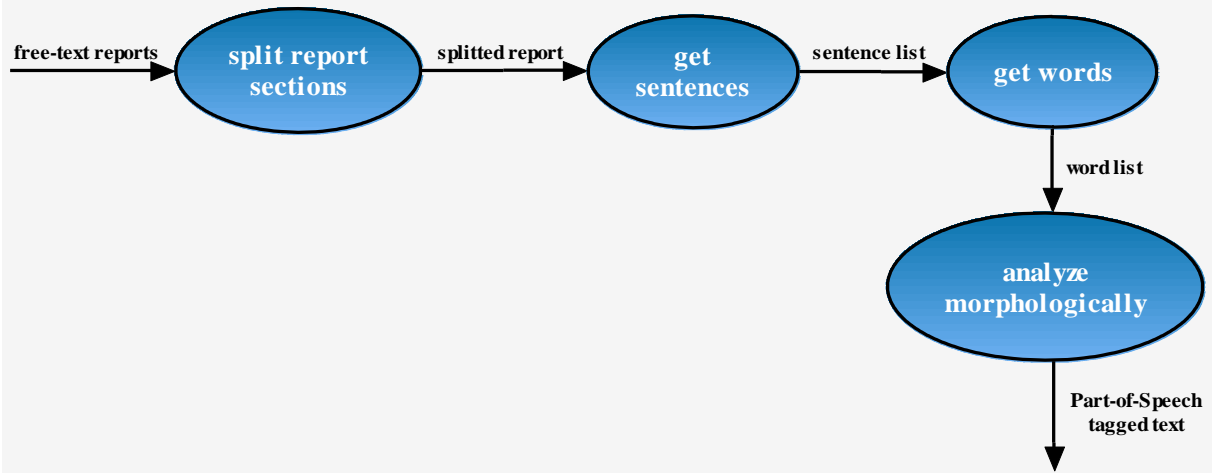
#### 4.2.1. DATA FLOW DIAGRAMS



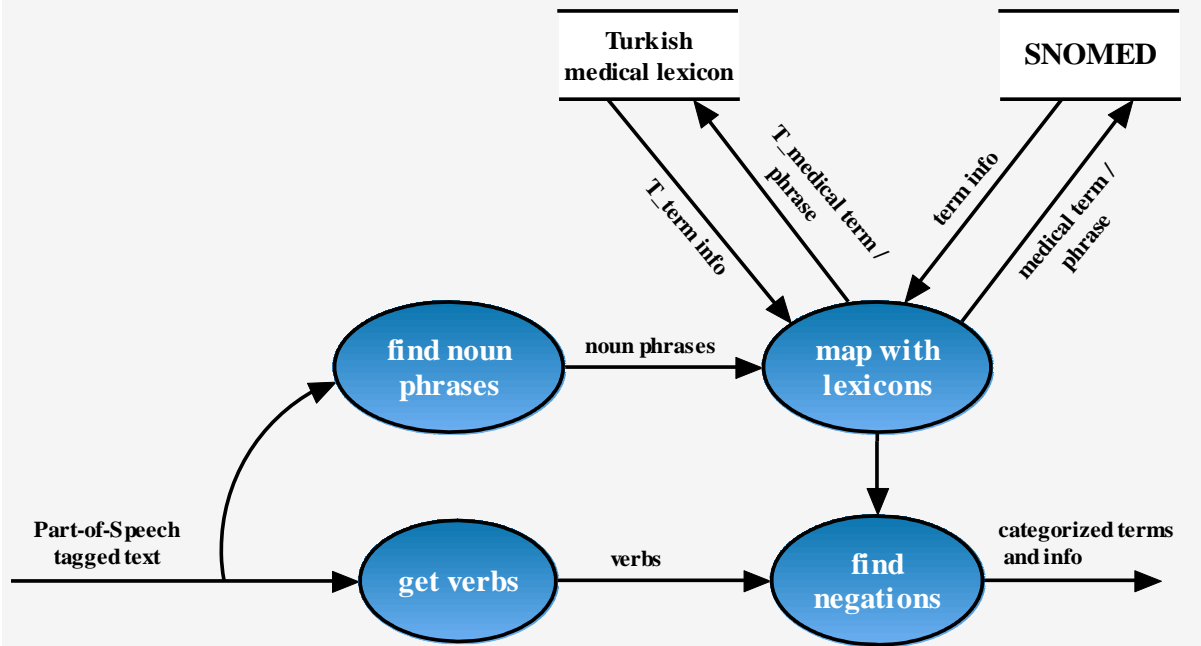
## Data Flow Diagram - LVL 1 "RadeX"



## Data Flow Diagram - LVL 2 "preprocess report"



## Data Flow Diagram - LVL 2 "associate semantics"





---

## 4.2.2. DATA FLOW EXPLANATION

---

### 4.2.2.1. PRE-PROCESSOR

---

This is the first part of the program. It takes the report as free text. Report will be divided into main sections which are apparent from the very structure of radiology reports:

- TUR
- Klinik Bilgi
- Teknik
- Bulgular
- Sonuc
- Doktorlar

Within each section, sentences will be separated and tagged. In sentences, words will be tagged separately.

In morphological analyze phase, words will be broken into roots and suffixes. Then, analyzer will tag the right POS associated with words.

We can use either Zemberek or Kemal Oflazer's Morphologic Analyzer (or may be both) to accomplish this process. Both of these two projects can find part of speech (POS) of words and also types of suffixes. But they may differ in types of suffixes.

Once we get POS of words and types of suffixes, they are ready to be sent to Semantic Association.

### 4.2.2.2. SEMANTIC ASSOCIATION

---

POS of words and types of suffixes arrive from "preprocess report". Term info about medical term arrives from SNOMED and Turkish medical lexicon (or other ontological medical glossaries).

Using these inputs, Semantic association deals with meaning of words or word phrases. Words are going to be examined in a single word base first. Then dependencies between words are going to be constructed according to their POS, suffixes and also meanings.

Program will use different medical glossaries to inspect whether a word is a medical term or not. Also, ontological relations between terms can be recognized with help of SNOMED and similar ontological glossaries. These relations will certainly contribute determining meanings of word phrases.

If program cannot find word (word phrase) in a medical (ontological) glossary and also in Turkish lexicon, i.e. the word is neither a Turkish word nor English term, program will try to translate the word into English using spell check tricks. If this does not work, it will try to "learn" whether it can be a term or not with learning techniques.

After getting meanings of words/ word phrases, this module puts the medical term in correct class. (i.e. whether it is "problem", "finding", "procedure", "device"), associate characteristics for terms. (e.g. For "problem": body location, certainty....) to form categorized terms and information related to them.

Semantic association finishes its job by sending categorized terms and info about them to Learning and Extraction Module.

#### 4.2.2.3. LEARN AND EXTRACT

---

Categorized terms and info will be analyzed here.

Discourse segmentations in a report will be figured out and dissolved in this part after analyzing categorized terms.

Applying various learning techniques to statistical data gained from all of the reports will result in meaningful information. The detail of this “meaningfulness” is up to which and how learning techniques applied efficiently.

The experience of the machine will be stored in Knowledge Base so that program will increase its correctness with new experiences.

This Module will send formatted and meaningful data to database so that it will be able to search and retrieve the data on demand.

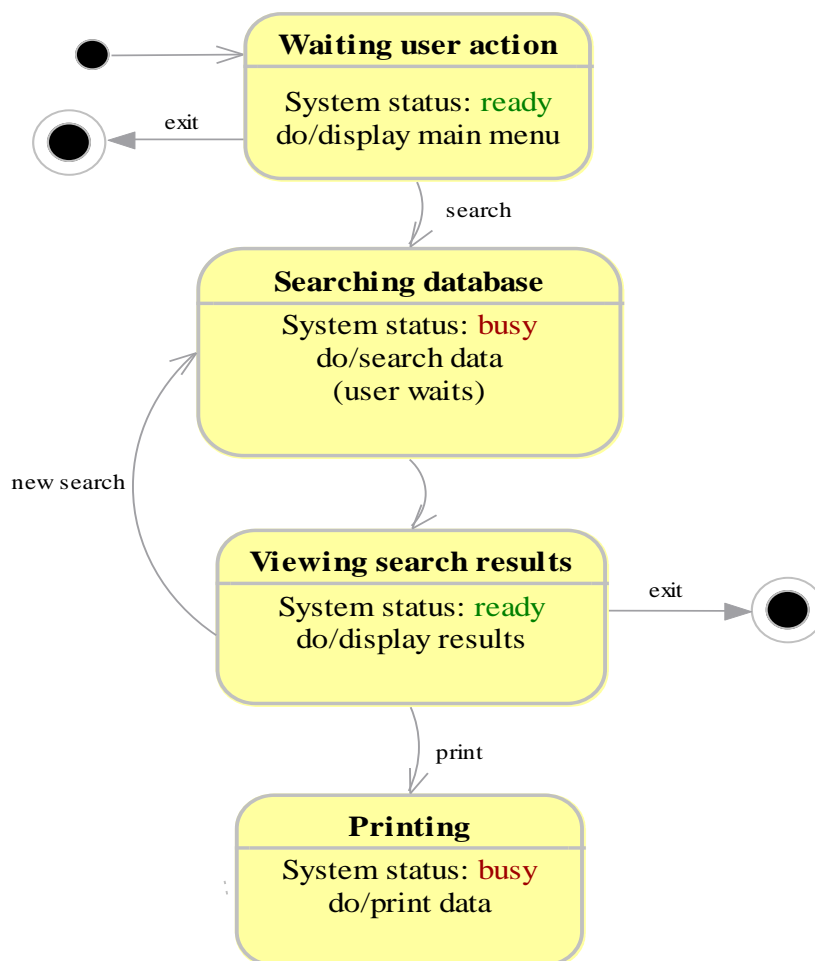
### 4.3. BEHAVIOURAL MODELING

---

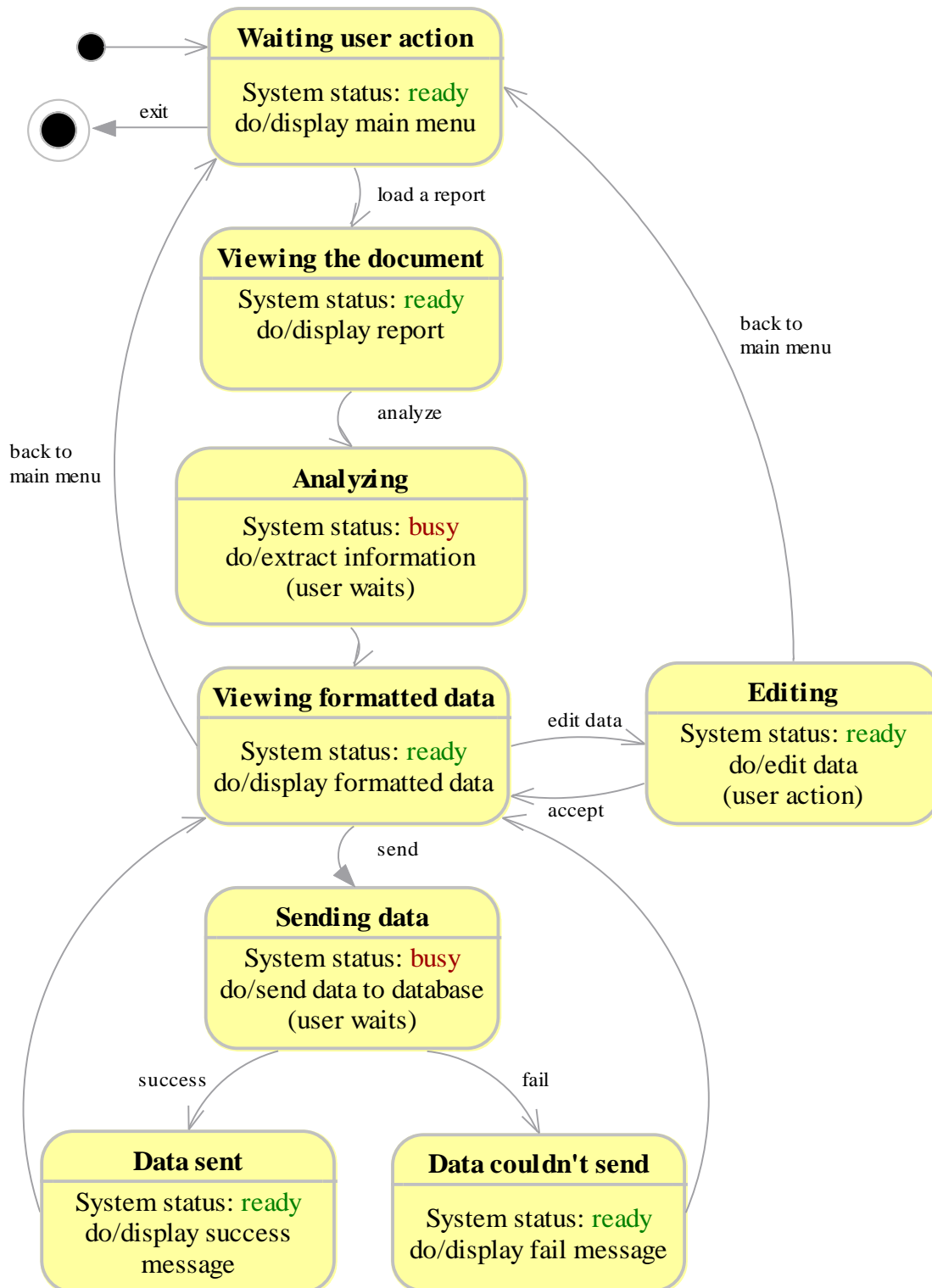
#### 4.3.1. STATE TRANSITION DIAGRAMS

---

Search Data - State Diagram



## Extract and send data - State Diagram



---

## 5. SCHEDULING

---

Our scheduling plan depicted by a Gantt chart can be seen at Appendix 7.2

---

## 6. PLANNING

---

---

### 6.1. RISK PLAN

---

Project's risks can be separated into three according to its sources which are design, group members and sponsor company.

**Design:** The most common mistake that we can encounter due to the process model, which we will use in our project, is that the several independently build pieces might not fit together. In order not to encounter such a problem, design will be carried out as detailed as possible. Despite our efforts, even if any mismatch occurs and cannot be resolved by the tester in a few days, every group member will pay attention to the problem and the problem will be fixed by the help of the design documentation. After that, the processes about the new model will continue.

**Group Members:** If any group member leaves the project for any reason, the responsibility of the person quitted will be shared among the left ones. There may be some conflicts on a subject among the members and a proper decision may not be made. In this case, the leader of the group will try to reconcile. But if this situation continues for two days, the leader will lead the group on his opinion.

**Sponsor Company:** If the communication with the sponsor company fails, source needs like medical reports can be provided by means of our own efforts (libraries, web and even hospitals).

---

### 6.2. QUALITY PLAN

---

Project's quality will be controlled from two perspectives.

The first one is business perspective which verifies whether the project could be completed up to final release deadline and whether it covers the needs of Sponsor Company.

The second one is the technical perspective that checks if the project conforms to user interface standards, documentation standards and naming standards. Moreover, successful integration of all modules, detailed testing and debugging should be done. The final implementation will comply with the final design report. The project should be well engineered in order to be robust and maintainable.

---

### 6.3. TEST PLAN

---

Since we are going to use incremental model in the project, model based testing is going to be used. After the completion of every build, functional test and performance testing will be applied. In the same time, studies of the new build will continue. Performance test will play the key role in developing a new build because the results encountered during testing in small builds will lead us with corrections and improvements for new builds. In this way, project will progress faster and more

stable. After completing the model progressions, it will be checked against software requirement specifications via validation tests.

## 6.4. EVALUATION AND VALIDATION METHOD

---

We plan to use ten-fold cross validation to evaluate the performance of the program: Reports are going to be divided into ten partitions. When nine partitions are used to train the system, the remaining partition will be used for testing. The testing will be repeated for all the other nine partitions as well. Then the results will be averaged to produce a single estimation. The performances for results will be measured according to *recall* and *precision*.

$$\text{Recall} = (\text{number of extracted true terms}) / (\text{number of total true terms})$$
$$\text{Precision} = (\text{number of extracted true terms}) / (\text{number of extracted terms})$$

Precision and recall for all single reports are averaged to find the precision and recall of the partition

## 7. APPENDIX

---

### 7.1. DATA DICTIONARY

---

**Name:** Free-text radiology reports

**Use area:** It's used as an input to the main program.

**Description:** Free-text radiology reports have a certain structure:

- Heading
- Clinical information
- Findings
- Conclusion

**Name:** Formatted data

**Use area:** It's shown to screen when the report is analyzed.

**Description:** It's the information extracted from the report. It's shown as XML format to the user.

**Name:** Query string

**Use area:** It's used as the query to search on database.

**Description:** It may be a simple keyword, a phrase or a group of keywords.

**Name:** Query result

**Use area:** It's shown to screen when a query is done and the database server responds the query successfully.

**Description:**

**Name:** Part of speech tagged text

**Use area:** It's used as input to the module "associate semantics".

**Description:** Every word in a POS tagged text carries POS information, i.e., whether it's a noun, adjective, verb etc. with it.

**Name:** Medical term / phrase

**Use area:** It's sent to SNOMED to by "associate semantics" module.

**Description:** It's possibly a medical term which has an ontological meaning in SNOMED.

**Name:** Term info

**Use area:** It's used to give semantics to a term.

**Description:** It's the information about the term like, whether it exists in SNOMED, its category and place in the ontological hierarchy.

**Name:** T\_Medical term / phrase

**Use area:** It's sent to Turkish medical lexicon by "associate semantics" module.

**Description:** It's possibly a medical term which has a medical meaning in the lexicon.

**Name:** T\_Term info

**Use area:** It's used to give semantics to a Turkish term.

**Description:** It's the information about the term like, whether it exists in the lexicon, its category and place in the ontological hierarchy.

**Name:** Categorized terms and info

**Use area:** It's the output of "associate semantics" module and the input for learning module.

**Description:** These are the words, to which some semantics are loaded by "associate semantics module"

**Name:** Current/New knowledge

**Use area:** It's used by the learning module.

**Description:** Current knowledge is stored in Knowledge-Base. As the learning algorithm generates new knowledge, current knowledge will be updated.

---

### 7.1.1. DATABASE TABLES

---

**Name:** Rapor

**Description:** This table holds the base information about an analyzed document, such as its heading and the date it was written.

**Name:** Islem

**Description:** This table holds information about a technique (ultrasound, chest x-ray, and radiograph) that subsist in the radiology report, or some treatment or examination that the patient received in the past or that is proposed in the radiology report.

**Name:** Problem

**Description:** This table holds detailed information about a medical problem that subsists in the document. Also it may be the case the patient doesn't suffer from the problem. We still hold information about it in this table.

**Name:** Normal\_bulgu

**Description:** This table holds detailed information about a normal condition finding, as depicted by the 'the heart size is normal'.

**Name:** Ilac\_tedavi

**Description:** This table holds information about a medication treatment that the patient received in the past, or one that is proposed in the radiology report.

**Name:** Oneri

**Description:** This table holds information about suggestions of some treatment, analysis or examination that is proposed in the analyzed radiology report.

## 7.2. GANTT CHART

**GANTT CHART - 9 MONTH TIME LINE**

