[TURKUAZ]

MIDDLE EAST TECHNICAL UNIVERSITY
DEPARTMENT OF COMPUTER ENGINEERING

‘Text Mining On Turkish Medical Radiology Reports’

USER MANUAL

By

Spring, 2007

Kerem Hadımlı – 1448752

Çiğdem Okuyucu – 1448976

Esra Zeynep Abacıoğlu – 1394568

Makbule Gülçin Özsoy – 1395383

İpek Tatlı – 1395557

Table of Contents

1. Introduction

1.1 What is *RadioRead* ?

RadioRead is a software product that extracts well-structured information from Turkish medical reports given as free text format. This product provides secure, ethical and user friendly access to patient information as easy as using a natural language. Our target audience can consist of doctors and statisticians who have little experience about computer usage. With our user friendly product, detailed information about patients can be accessed easily; more information about a patient can be given to his/her doctor before consultations; the information can be used by doctors to diagnose diseases of other patients; and statistics can be derived with no need of any technical knowledge.

1.2 Requirements

- PostgreSQL Database Management System
- Java Run Time Environment 6
- Windows XP or recent GNU/Linux Distribution
- Internet access for online dictionary support (Zargan, TDK)

1.3 Installation

To install RadioRead, please use the radioread-install.jar file. You can double click on this file on Windows, and you can start the file manually by typing `java -jar radioread-install.jar` on Unix / Linux systems.

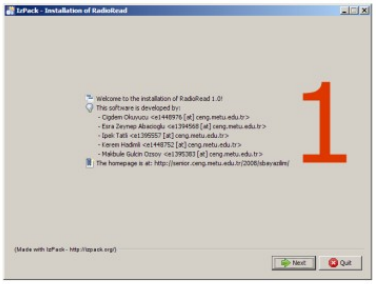
After showing the readme (*2nd screen*), the installer will ask the path to install RadioRead (*3rd screen*). After this step, you have to choose the packages you want to install (*4th screen*). You can choose to install *Documentation* and *SQL Initialization* packages besides the *Base* package.

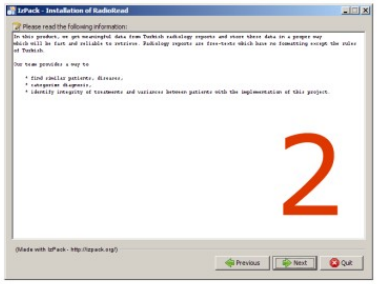
During the installation, you have to provide the installer with the database connection credentials (*5th screen*). A PostgreSQL database server needs to be present that is accessible over the network (or on local computer). The installation program will guide you through the steps. After installation is complete, the SQL to initialize the database will be shown (*7th screen*).

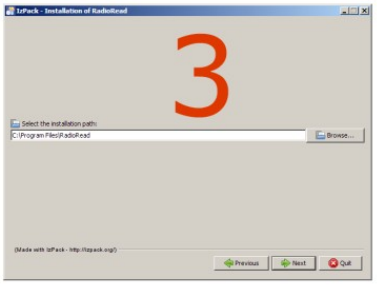
If you want to install RadioRead to a diverse organization on multiple computers, you may choose to create an automated installer that remembers your settings at the last step (*8th screen*).

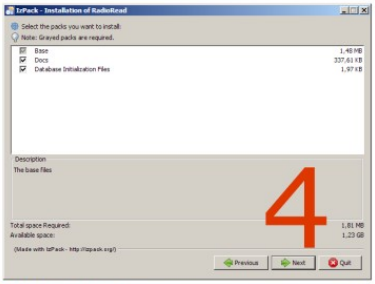
The SQL file to initialize the database (which is shown in *7th screen*) can also be found in <*installation directory*>/SQL/tables.sql file if you select Docs package while installing (*4th screen*).

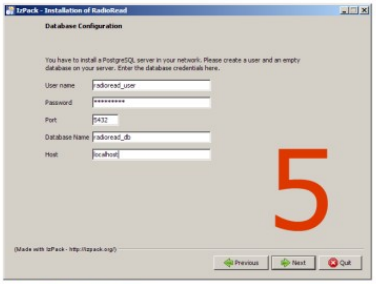
To start radioread, double click on <*installation directory*>/RadioRead.jar. Please make sure you applied the SQL initialization file before starting RadioRead.

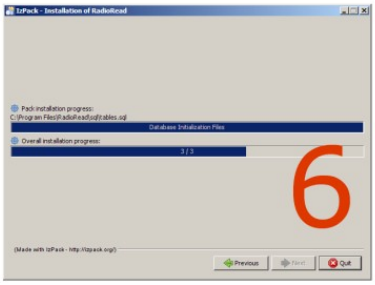

IzPack - Installation of RadioRead
Welcome to the installation of RadioRead 1.0!
This software is developed by:
- Cigdem Okuyucu <ei1449576 [at] ceng.metu.edu.tr>
- Esra Zeynep Abacioglu <ei1394568 [at] ceng.metu.edu.tr>
- Ipek Tatli <ei1399557 [at] ceng.metu.edu.tr>
- Kerem Hadimli <ei1449752 [at] ceng.metu.edu.tr>
- Makbule Gulcin Ozsoy <ei1395383 [at] ceng.metu.edu.tr>
The homepage is at: http://senior.ceng.metu.edu.tr/2008/sbavazilim/
(Made with IzPack - http://izpack.org/)
Next
Quit
1

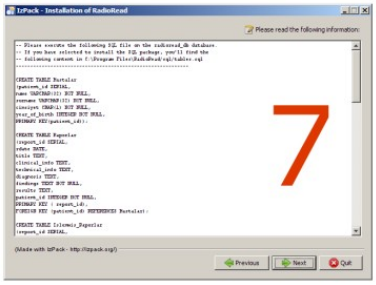

IzPack - Installation of RadioRead
Please read the following information:
Our team provides a way to
* find similar patients, diseases,
* categorize diagnosis,
* identify integrity of treatments and variances between patients with the implementation of this project.
(Made with IzPack - http://izpack.org/)
Previous
Next
Quit
2

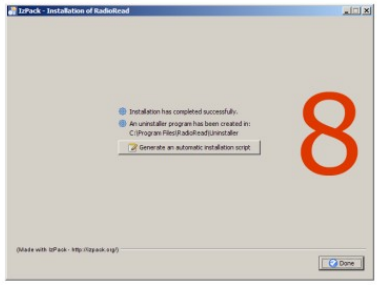

IzPack - Installation of RadioRead
3
Select the installation path:
C:\Program Files\RadioRead
Browse...
(Made with IzPack - http://izpack.org/)
Previous
Next
Quit


IzPack - Installation of RadioRead
Select the packs you want to install.
Note: Grayed packs are required.
Base 1.48 MB
Docs 337.61 KB
Database Initialization Files 1.97 KB
Description
The base files
Total space Required: 1.81 MB
Available space: 1.23 GB
(Made with IzPack - http://izpack.org/)
Previous
Next
Quit
4


IzPack - Installation of RadioRead
Database Configuration
You have to install a PostgreSQL server in your network. Please create a user and an empty database on your server. Enter the database credentials here.
User name radoread_user
Password
Port 5432
Database Name radoread_db
Host localhost
(Made with IzPack - http://izpack.org/)
Previous
Next
Quit
5


IzPack - Installation of RadioRead
Pack installation progress:
C:\Program Files\RadioRead\sql\tables.sql
Database Initialization Files
Overall installation progress:
3 / 3
(Made with IzPack - http://izpack.org/)
Previous
Next
Quit
6


IzPack - Installation of RadioRead
Please read the following information:
-- Please execute the following SQL file on the radioread_db database.
-- If you have selected to install the SQL package, you'll find the
-- following content in C:\Program Files\RadioRead\sql\tables.sql
CREATE TABLE Kurallar
(patient_id SERIAL,
name VARCHAR(10) NOT NULL,
surname VARCHAR(15) NOT NULL,
sinsiyet CHAR(1) NOT NULL,
year_of_birth INTEGER NOT NULL,
PRIMARY KEY(patient_id));
CREATE TABLE Raporlar
(report_id SERIAL,
mdate DATE,
title TEXT,
clinical_info TEXT,
technical_info TEXT,
diagnosis TEXT,
findings TEXT NOT NULL,
results TEXT,
patient_id INTEGER NOT NULL,
PRIMARY KEY (report_id),
FOREIGN KEY (patient_id) REFERENCES Kurallar);
CREATE TABLE Islemler_Raporlar
(report_id SERIAL,
(Made with IzPack - http://izpack.org/)
Previous
Next
Quit
7

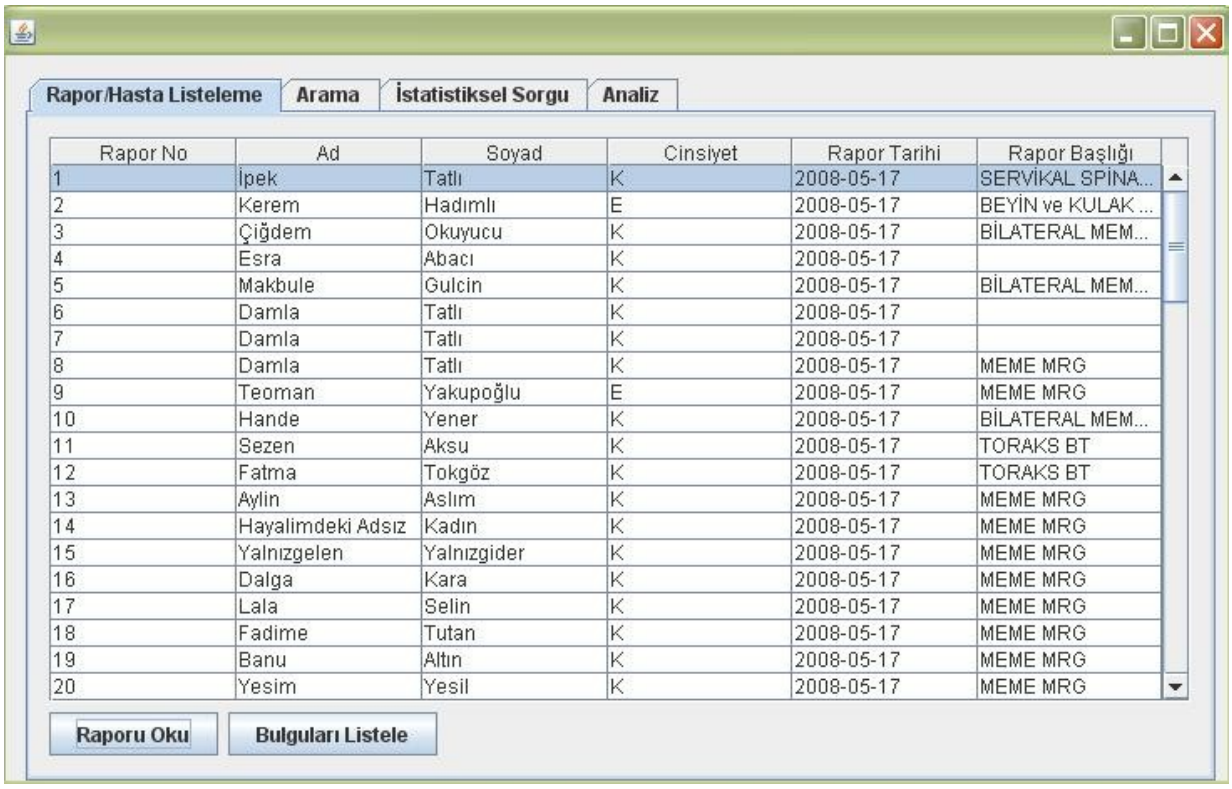

IzPack - Installation of RadioRead
Installation has completed successfully.
An uninstaller program has been created in:
C:\Program Files\RadioRead\Uninstaller
Generate an automatic installation script
(Made with IzPack - http://izpack.org/)
Done
8

2. RadioRead Window

Our main window consists of four tabs which are namely List of Report/Patient (Rapor / Hasta Listeleme) , Search (Arama), Statistical Queries (Istatistiksel Sorgu), Analyze (Analiz).

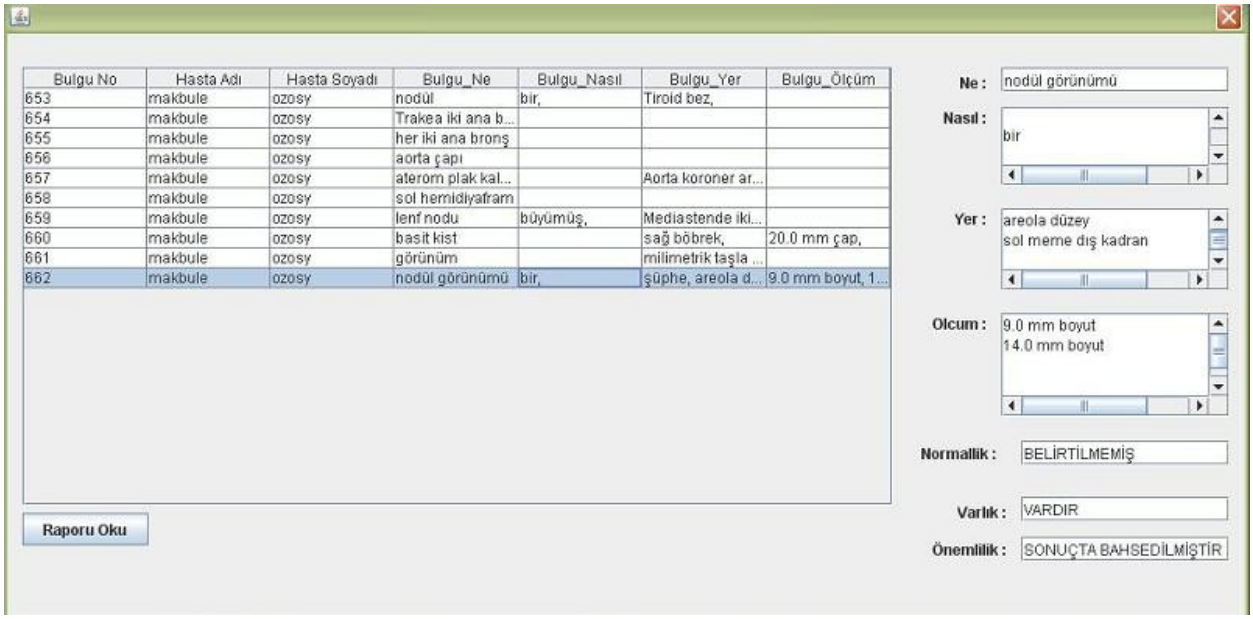
2.1 List of Report/Patient (Rapor / Hasta Listeleme)

In this part, information about the users and the reports can be accessed. The information consist of report id , name, surname , gender, date of report and header of report. In addition to get information, user can also access the original report by clicking the Read Report (Raporu Oku) button and the information about the chosen finding by clicking List Finding (Bulguları Listele) button. When the Read Report (Raporu Oku) button is clicked a new screen pops up, then user can access the finding list by clicking the List Finding (Bulguları Listele) button. When the List Finding (Bulguları Listele) button is clicked a new screen pops up. In this new screen when the user selects a row; the information in that row (namely; what, quality, location, measurement, normality, existency and importance) can be seen in the right part of the screen.



Rapor No	Ad	Soyad	Cinsiyet	Rapor Tarihi	Rapor Başlığı
1	İpek	Tatlı	K	2008-05-17	SERVİKAL SPİNA...
2	Kerem	Hadımlı	E	2008-05-17	BEYİN ve KULAK ...
3	Çiğdem	Okuyucu	K	2008-05-17	BİLATERAL MEM...
4	Esra	Abacı	K	2008-05-17	
5	Makbule	Gulcin	K	2008-05-17	BİLATERAL MEM...
6	Damla	Tatlı	K	2008-05-17	
7	Damla	Tatlı	K	2008-05-17	
8	Damla	Tatlı	K	2008-05-17	MEME MRG
9	Teoman	Yakupoğlu	E	2008-05-17	MEME MRG
10	Hande	Yener	K	2008-05-17	BİLATERAL MEM...
11	Sezen	Aksu	K	2008-05-17	TORAKS BT
12	Fatma	Tokgöz	K	2008-05-17	TORAKS BT
13	Aylin	Aslım	K	2008-05-17	MEME MRG
14	Hayalimdeki Adsız	Kadın	K	2008-05-17	MEME MRG
15	Yalnızgelen	Yalnızgider	K	2008-05-17	MEME MRG
16	Dalga	Kara	K	2008-05-17	MEME MRG
17	Lala	Selin	K	2008-05-17	MEME MRG
18	Fadime	Tutan	K	2008-05-17	MEME MRG
19	Banu	Altın	K	2008-05-17	MEME MRG
20	Yesim	Yesil	K	2008-05-17	MEME MRG

Figure 1. List of Report / Patient (Rapor / Hasta Listeleme) Tab



Bulgu No	Hasta Adı	Hasta Soyadı	Bulgu_Ne	Bulgu_Nasıl	Bulgu_Yer	Bulgu_Ölçüm
653	makbule	ozosy	nodül	bir,	Tiroid bez,	
654	makbule	ozosy	Trakea iki ana b...			
655	makbule	ozosy	her iki ana bronş			
656	makbule	ozosy	aorta çapı			
657	makbule	ozosy	aterom plak kal...		Aorta koroner ar...	
658	makbule	ozosy	sol hemidiyafram			
659	makbule	ozosy	lenf nodu	büyümüş,	Mediastende iki...	
660	makbule	ozosy	basit kist		sağ böbrek,	20.0 mm çap,
661	makbule	ozosy	görünüm		milimetrik taşla ...	
662	makbule	ozosy	nodül görünümü	bir,	şüphe, areola d...	9.0 mm boyut, 1...

Figure 2. Interface after clicking on “Bulguları Listele” Button

Figure 3. Interface after clicking on “Raporu Oku” Button

2.2 Search (Arama)

There are three parts in this tab, namely Search Finding (Bulgu Arama), Search Patient (Hasta Arama), Search Report (Rapor Arama). In the usage of Search (Arama) tab, the user will see several punctuations, they help the user to understand the search criteria. These punctuations are [ ] user fills this blank by inserting finding information, { } user fills this blank by inserting report information, ( ) user fills this blank by inserting location

information. [...] indicates that, this blank can be filled by the type of the nearest punctuation. Every part has a capability to edit the selected text.

Firstly, Search Finding (Bulgu Arama) tab is selected, and then the user creates his/her own search criterias using selection links. When the link is clicked, user can create his/her own criterias by selecting the appropriated boxes in the tree. The user chooses location information as criteria and the search will be done after the user clicked Search (Ara) button. The result screen comes as a pop up, here the report and the findings can be seen according to search criteria of the user. From the pop up user can access the original report. The user can change his/her search criteria by clicking the Clean (Temizle) button.

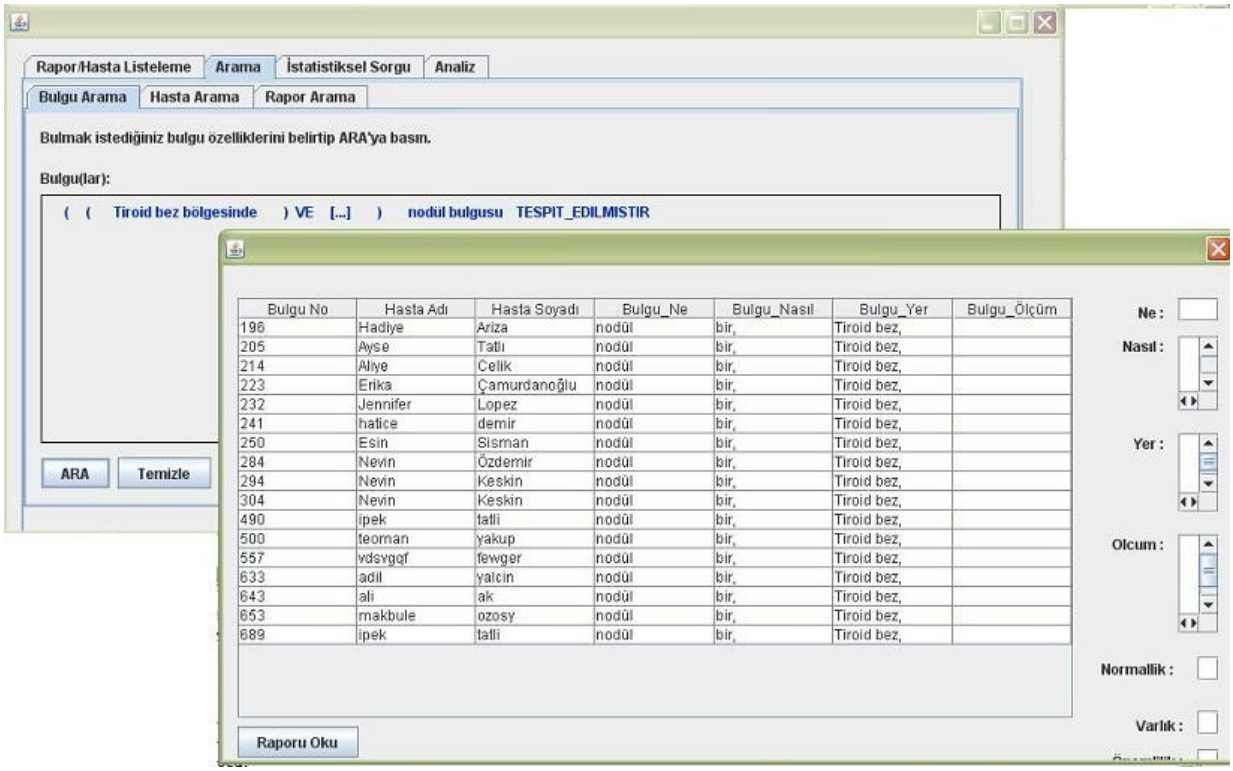


Figure 4. Finding Search (Bulgu Arama) Tab of Search (Arama) Tab

Secondly, Search Patient (Hasta Arama) tab is selected, and then the user creates his/her own search criterias using selection links. When the link is clicked, user can create his/her own criterias by selecting the appropriated boxes in the tree. The user chooses any combinations of these information: age range, gender, date of the report, findings, location as criteria and the search will be done after the user clicked Search (Ara) button. The result

screen comes as a pop up; here the report and the findings can be seen according to search criteria of the user. From the pop up user can access the original report and the all findings of the patient. The user can change his/her search criteria by clicking the Clean (Temizle) button.

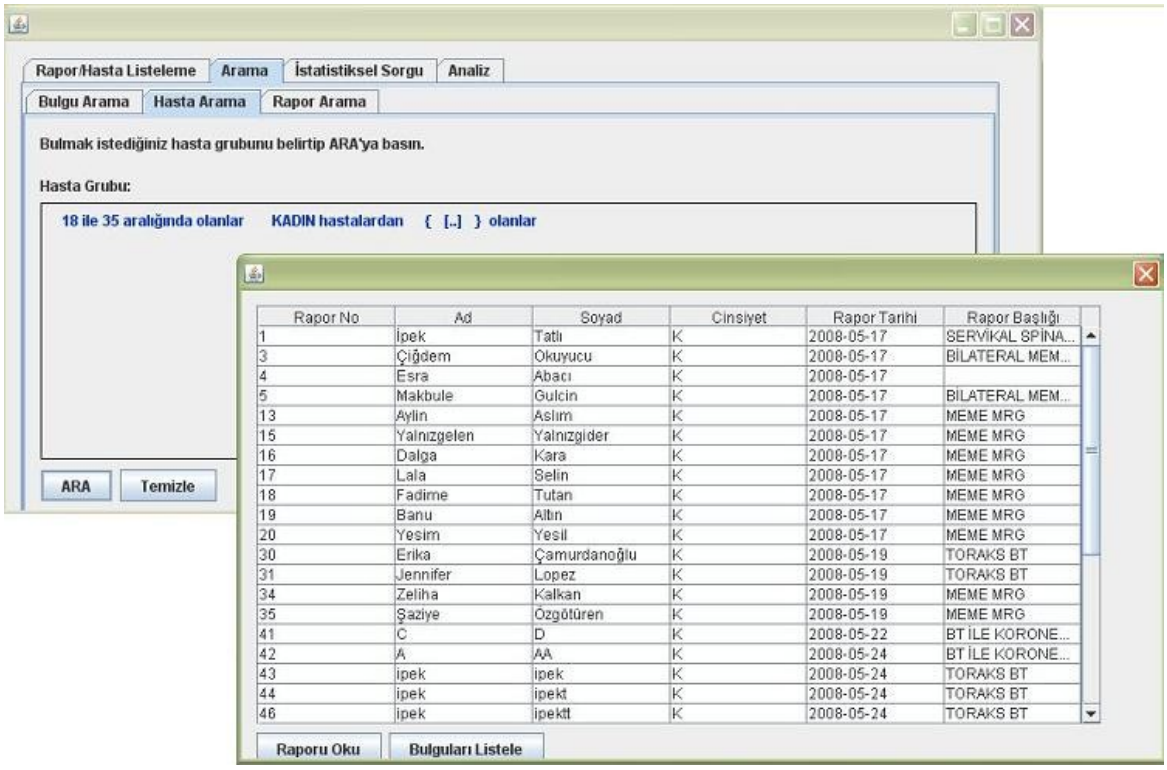


Figure 5. Patient Search (Hasta Arama) Tab of Search (Arama) Tab

Thirdly, Search Report (Rapor Arama) tab is selected, and then the user creates his/her own search criterias using selection links. When the link is clicked, user can create his/her own criterias by selecting the appropriated boxes in the tree. The user chooses any combinations of these information: what, quality, location, measurement, normality, existency, importance as criteria and the search will be done after the user clicked Search (Ara) button. The result screen comes as a pop up, here the report and the findings can be seen according to search criteria of the user. From the pop up user can access the original report and the all findings of the patient. The user can change his/her search criteria by clicking the Clean (Temizle) button.

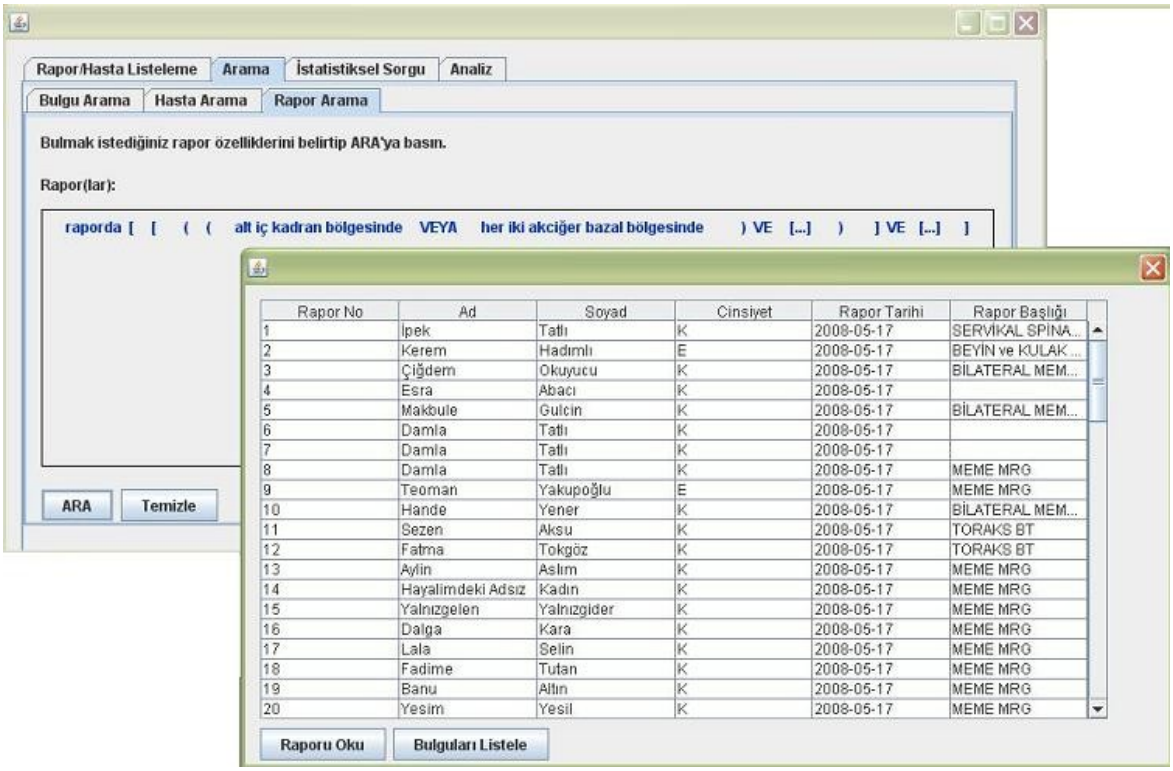


Figure 6. Report Search (Repor Arama) Tab of Search (Arama) Tab

2.3 Statistical Queries (Istatistiksel Sorgu)

There are three parts in this tab, namely Number Of Patient (Hasta Sayısı) ,Chart (Cizelge) , Ratio Of Patient (Hasta Oranı). In the usage of Statistical Queries (Istatistiksel Sorgu) tab, the user will see several punctuations, they help the user to understand the search criteria. These punctuations are [ ] user fills this blank by inserting finding information, { } user fills this blank by inserting report information, ( ) user fills this blank by inserting location information. [...] indicates that, this blank can be filled by the type of the nearest punctuation.

Firstly, in the Number Of Patient (Hasta Sayısı) tab, user can edit the criteria of the query by clicking on the editable areas, these areas consist of the report information. User can select any of the report information as criteria of query. After clicking on the Calculate (Hesapla) button, user can see the number of the patient that fits with the query. The user can change his/her criteria of the query by clicking the Clean (Temizle) button.

Figure 7. Number of Patient (Hasta Sayısı) Tab of Statistical Queries (Istatistiksel Sorgu) Tab

Secondly, in the Chart (Cizelge) tab, user can edit three part. First of them is in the group of Patient Part (Hasta Grubu), user can edit report information according to his/her criteria. Second of them is in the Finding (Bulgu), user can edit any location information of the finding. Third of them is in the Column (Sutun), user can enter the column number which will be the range number in the result. The user can change his/her criteria by clicking the Clean (Temizle) button.

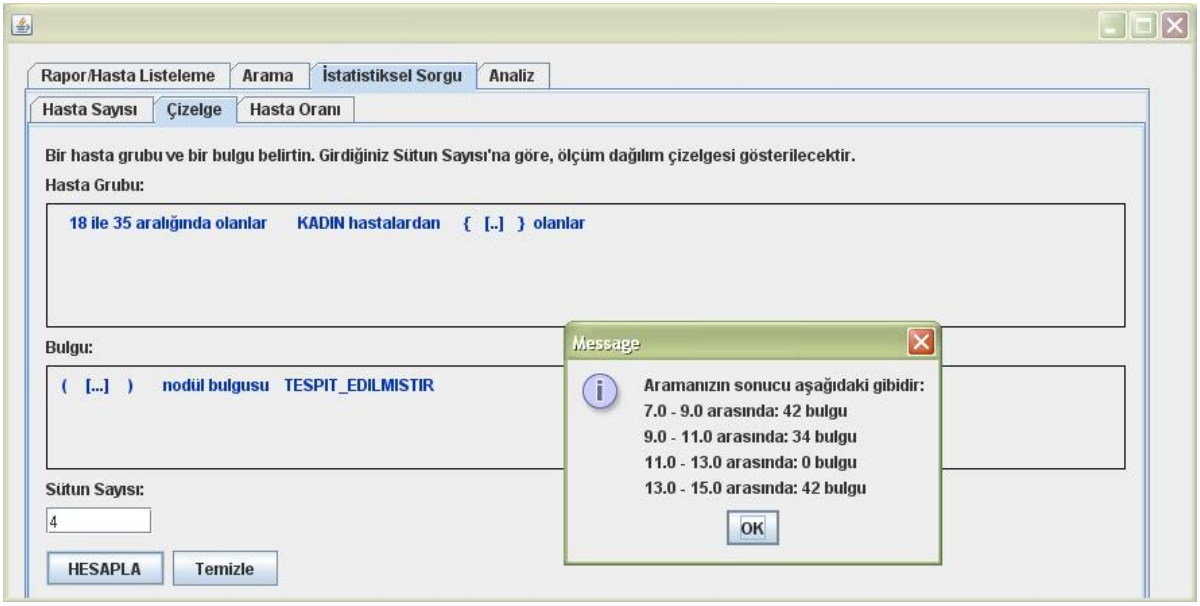


Figure 8. Chart (Çizelge) Tab of Statistical Queries (Istatistiksel Sorgu) Tab

Thirdly, in the Ratio Of Patient (Hasta Oranı) tab, user can edit two main part: Patient Group A (Hasta Grubu A) and Patient Group B (Hasta Grubu B) with report information. Patient A group should be subset of the Patient B group. After clicking on the Calculate (Hesapla) button, user gets the number of the patient in the Patient Group A (Hasta Grubu A), the number of the patient in the Patient Group B (Hasta Grubu B) and finally the percentage of the number of the patient in the Patient Group A (Hasta Grubu A) / the number of the patient in the Patient Group B (Hasta Grubu B). The user can change his/her criteria by clicking the Clean (Temizle) button.

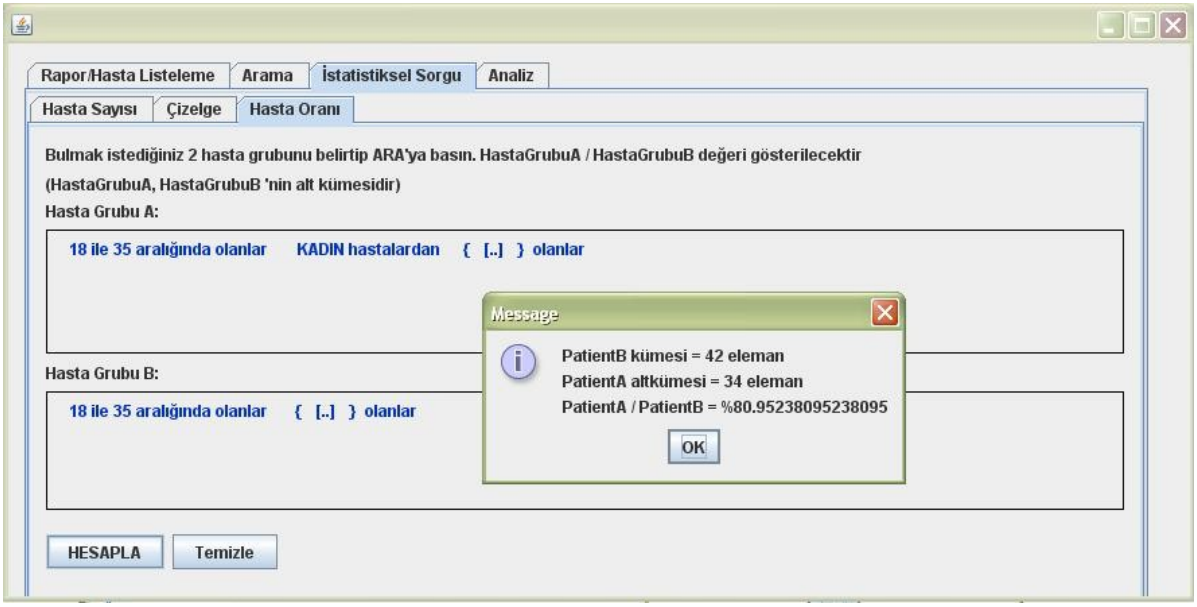


Figure 9. Patient Ratio (Hasta Oranı) Tab of Statistical Queries

(Istatistiksel Sorgu) Tab

2.4 Analyze (Analiz)

In this tab, thanks to the Load Report (Dosyadan Rapor Yukle), user can select any report from the hard disk. Then, edit the patient information, namely: name, surname, gender, year of birth. After clicking the Analyze Report (Rapor Analiz Et) button, a new pop up screen is created to get information form the user on medical terms. Finally, analyzed report is

inserted the database with the given patient information and doctor feedback, all these information can be seen on a new pop up screen.

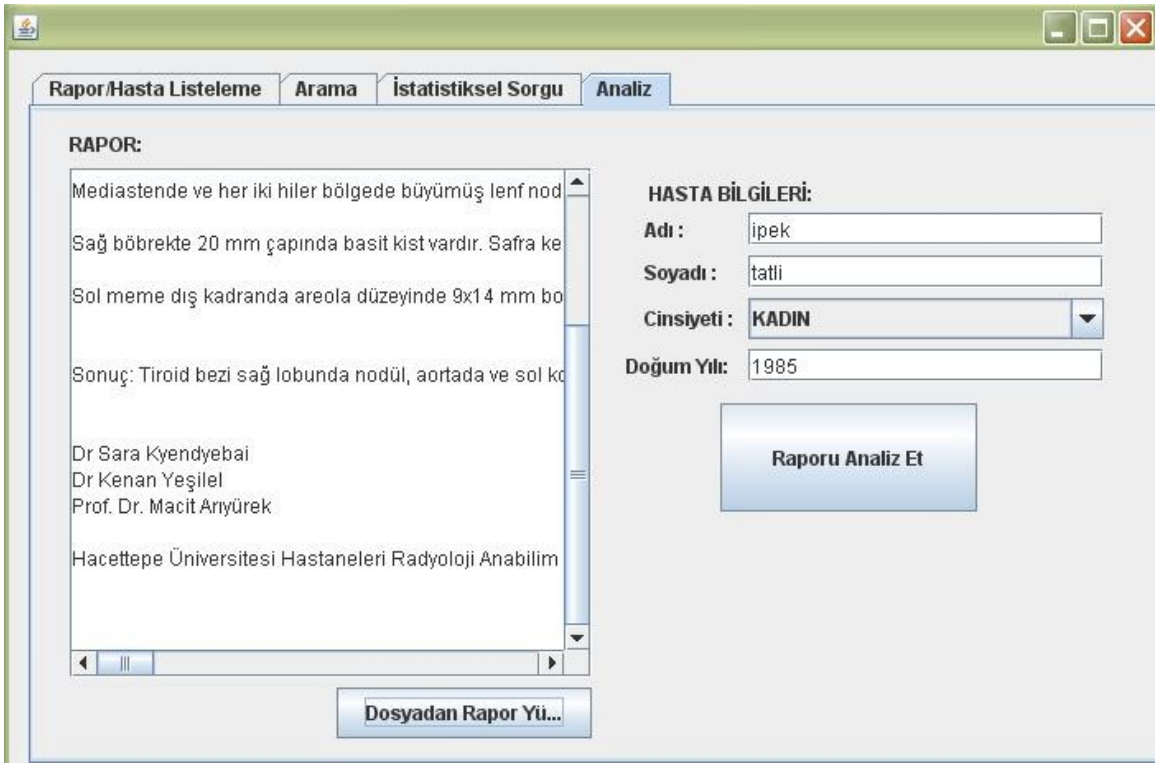


Figure 10. Analyze (Analiz) Tab

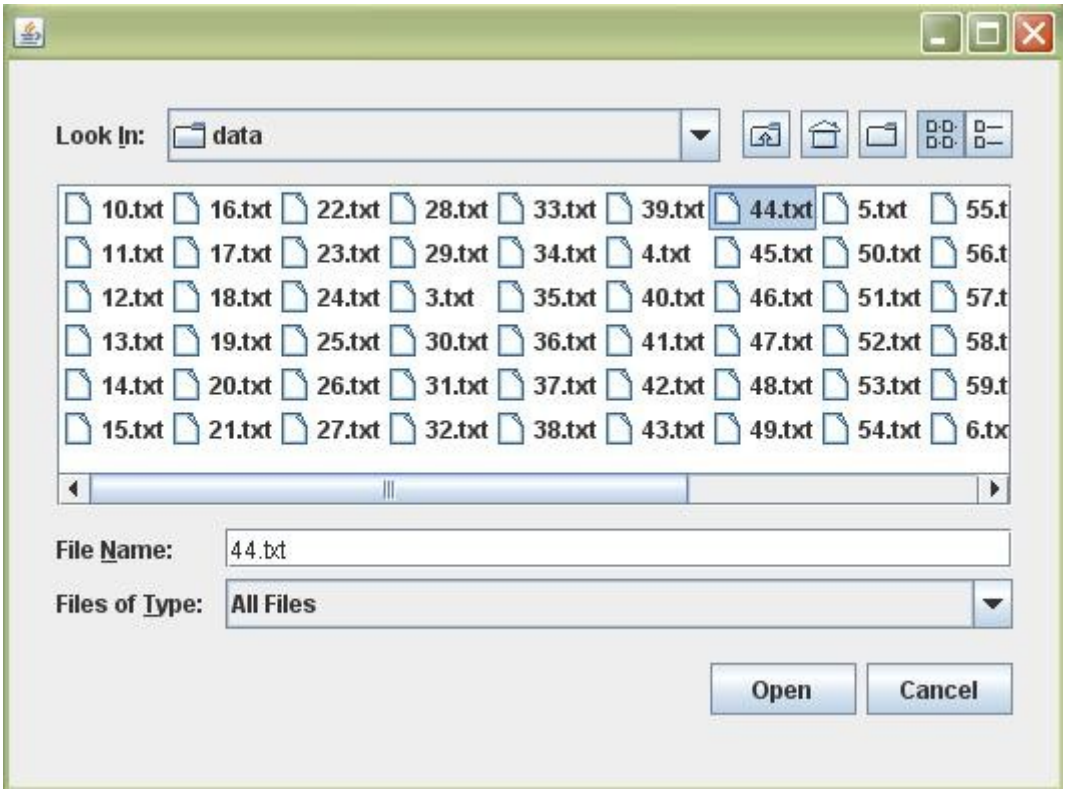


Figure 11. Browsing Report From File (Dosyadan Rapor Yükle)

3. Software Specification

RadioRead is distributed as a self-installing JAR file. Everything it requires, except for database is included.

3.1. Report File Format

RadioRead expects input report files to be in a near-strict format. There are 4 available sections, which may be or not be present. The following is a sample report which contains all 4 sections. Reports are required to have a title in their first lines.

```
TORAKS BT

Klinik bilgi: Sol akciğerde kitle.

Teknik: İVKM sonrası 5 mm kalınlığında transvers kesitler alınmıştır.

Bulgular: Trakea ve her iki ana bronş normaldir. Mediastende ve her
iki hiler bölgede büyümüş lenf nodu yoktur. Mediastinal ana vasküler
yapıların çapı, kalp büyüklüğü normaldir.

      Her iki akciğerde üst lob apikal bölgede belirgin olan
paraseptal amfizem vardır.

      Sağ akciğerde benign özellikte birkaç milimetrik kalsifiye nodül
vardır.

Sonuç: Sol akciğer alt lob süperior segmentte içerisinde hava
bronkogramı bulunan düzensiz konturlu nodül (akciğer karsinomu?), her
iki akciğerde buzlu cam dansitesinde nodüller.

Dr. XXXXX XXXXX
Prof. Dr. XXXXXX XXXXX

Hacettepe Üniversitesi Hastaneleri Radyoloji Anabilim Dalı'nın
radyolojik inceleme raporudur.
```

3.2. Internet Connection details

RadioRead requires an active internet connection while analyzing reports. This behaviour is a requirement for finding roots of words that internal library (Zemberek) cannot

find. RadioRead access http://www.zargan.com dictionary for root look-up. The results are cached in database.

3.3. Morphological Analyzer Limitations

RadioRead uses Zemberek library as its morphological analyzer, with external additions like online-dictionary check. Some words and inflected words (“çekilmiş sözcük”) are known to have problems with Zemberek library.

3.4. Noun Phrase Parser Limitations

RadioRead uses its own library for parsing complex noun phrases in Turkish. Although the library can parse most of noun phrases, and especially most of noun phrases used in radiology reports, the noun phrase parser has its own limitations. For example phrases like “Mide ve beyin kenarı” may lead to results like “Mide kenarı” “Beyin kenarı”, although this may not be what is meant.

Radioread asks questions according to meaning to the user. There should be a doctor present while analyzing reports, as important questions about terminology and qualifiers may be asked to the operator.

3.5. Database

RadioRead uses PostgreSQL as its database. PostgreSQL needs to be present in the system or network in order to use RadioRead. While other SQL servers might work, they won’t receive support by SBAYazilim.

The SQL file to initialize the database can be found in the installation folder, in the SQL directory.

RadioRead's SQL Server connection details can be found in installation directory, with the name "database.ini".