EHR Tools Summary

EHR application has seven main tools: XML generate tool, JSON generate tool, Database import tool, Machine learning data processing tool, query tools, statistics tools and prediction tools. Figure 1 show the architecture of application. Struts 2, Spring and Hibernate (SSH) make up a framework for development of Java web application. It uses the Model-View-Controller design pattern and JavaBean as the basic technologies. Java and JSP are used as the implementing language in SSH development.Above this framework, we built the application with different functions, including query, statistics and prediction function. MongoDB is the basic database to store records.

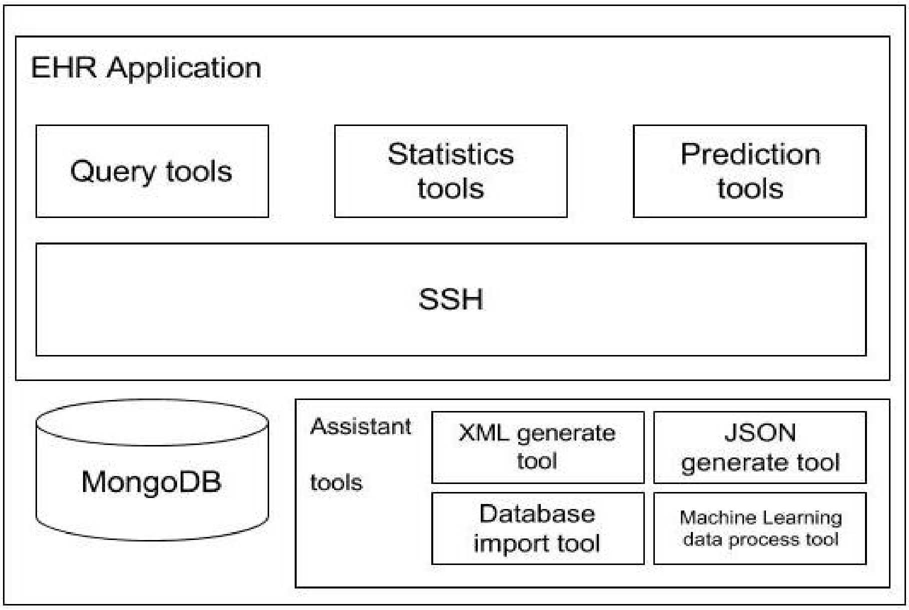


Figure 1. The architecture of EHR application

Tool 1:   XML file generation tool

Input:	The directory of .rtf data files stored(d:\\data)
Output	The XML data files in a specified directory (d:\\xml)
Function	Convert .rtf data files to the XML data files with specified directory.
API	method: readAndWriteXML() input:

	src ———- The directory of .rtf data files stored dst ——— The XML data files in a specified directory output: The XML data files
Code File:	Java Project: EhealthInfo Code file: RTFToXML.java

1. The data files of records

There are four groups of records with different years, include 2009, 2010, 2011 and 2012. All records are kept in text files with .rtf as a suffix.

2. The keywords of records

The key words of patient's records are batches (or years), registration number of records, hospital, department, date, patient's basic information (name, age, gender, profession, phone number, contact and address), symptoms, the diagnosis of western and Chinese medicine, western medicines, traditional Chinese medicines, process and the doctor.

3. The structure of XML files

The XML has a standard and unified structure based on the key words of records. You can find the XML structure in appendix 1.

Tool 2: JSON file generation tool

Input:	The directory of XML data files (d:\\ xml)
Output:	The JSON data files with specified directory(d:\\json)
Function:	Convert those XML data files to the JSON data files with specified directory to store in MongoDB.
API:	method: creatrJSONFile() input: src ——— The directory of XML data files stored dst ——— The JSON data files in a specified directory output: The JSON data files
Code File:	Java project: EhealthInfo Code file: XMLToJSON.java

We use the MongoDB, a document-oriented database, as the database to read and write data easier and faster. All data in MongoDB are stored into BSON files, which is similar to JSON format. So records should be converted to JSON files from XML files at first and then can be imported into MongoDB. The structure of JSON is same with XML files only with different grammar.

Tool 3: Database import tool

Input:	The directory of JSON data files
Output:	True or False
Function:	Import those JSON data files to Mongoldb
Code File:	No code file. Use the MongoDB import command: mongoimport -d <db name> -c <collection name> file.json

MongoDB is a cross-platform document-oriented database. Classified as a NoSQL database, MongoDB eschews the traditional table-based relational database structure in favor of JSON-like documents with dynamic schemas (MongoDB calls the format BSON), making the integration of data in certain types of applications easier and faster. With those JSON data files generated by Tool two, we can use the MongoDB tools to import those JSON-format data to MongoDB.

Tool 4: Data Query tools

1. Get records based on patient's name, process method and traditional Chinese medicines.(Figure 2 and Figure 3)

Input:	Batch or patient name or process method or traditional Chinese medicines
Output:	Patient's records information
Function:	Get records based on patient's name, process methods, and traditional Chinese medicines
API:	method: queryRecordsByCondition() input: batch ———- the batch(year) of records pname ——— patient's name process ——— process methods medicines ———— transitional Chinese medicines output: The list of target records
Code File:	Java project: EHR Code file: 1. findRecordsByCondition.jsp

	2. QueryAction.java

2. Get records based on the condition descriptions.(Figure 4)

Input:	Condition descriptions
Output:	Patient's records information
Function:	Get records based on the condition descriptions.
API:	method: queryRecordByInput() input: diagnose ———- the diagnose of records description ———- condition descriptions output: The list of target records
Code File:	Java project: EHR Code file: 1. findRecordByInputDescription.jsp 2. QueryAction.java

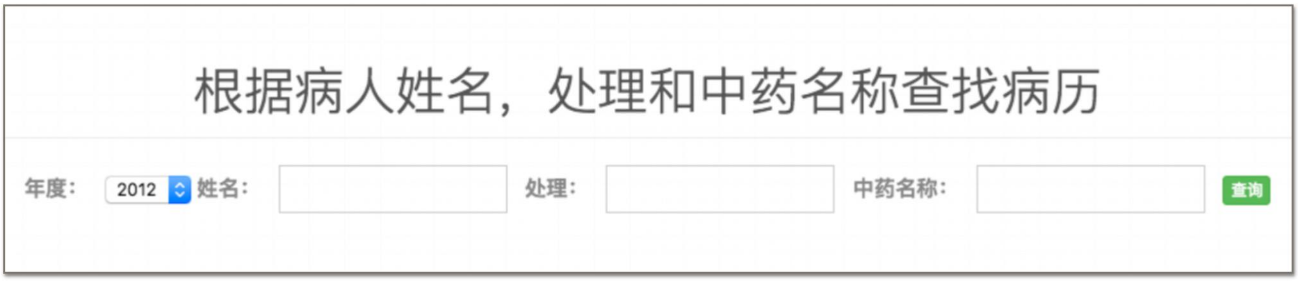
3. Get the details of record based the order and registration number of records

Input:	Registration number or order number of records
Output:	Patient's records information
Function:	Get the details of record based the order number or registration number of records.
API:	method: queryRecordByNo() input: String: count ———registration number or order number of records output: EhealthRecord: targetRecord———— the target record based on input
Code File:	Java project: EHR Code file:

	1. recordpreview.jsp 2. QueryAction.java

1. Keeping the privacy of records

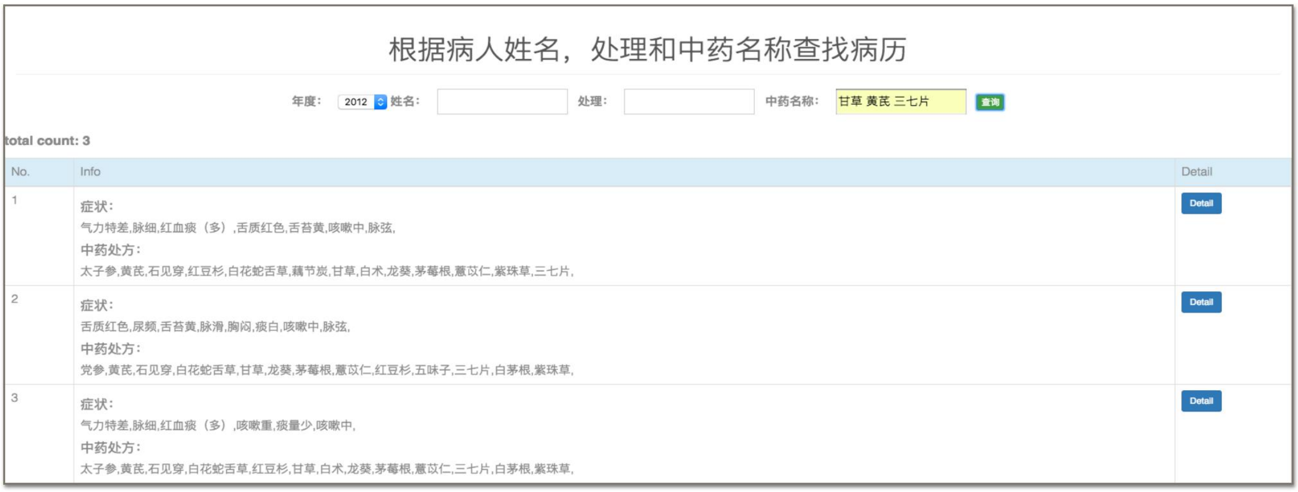
We need to take care of the privacy of patient record information by encrypting patient's basic information. The patient's basic information are name, age, gender, profession, phone number, contact and address. Whenever we query the records, the basic information would be encrypted automatically and be replaced by the encrypted string except patient's name. The name of the patient should be processed specifically by keeping first name but last name being replaced by encrypted strings.



根据病人姓名，处理和中药名称查找病历

年度： 2012 姓名： 处理： 中药名称：

Figure 2. main page of getting records based conditions



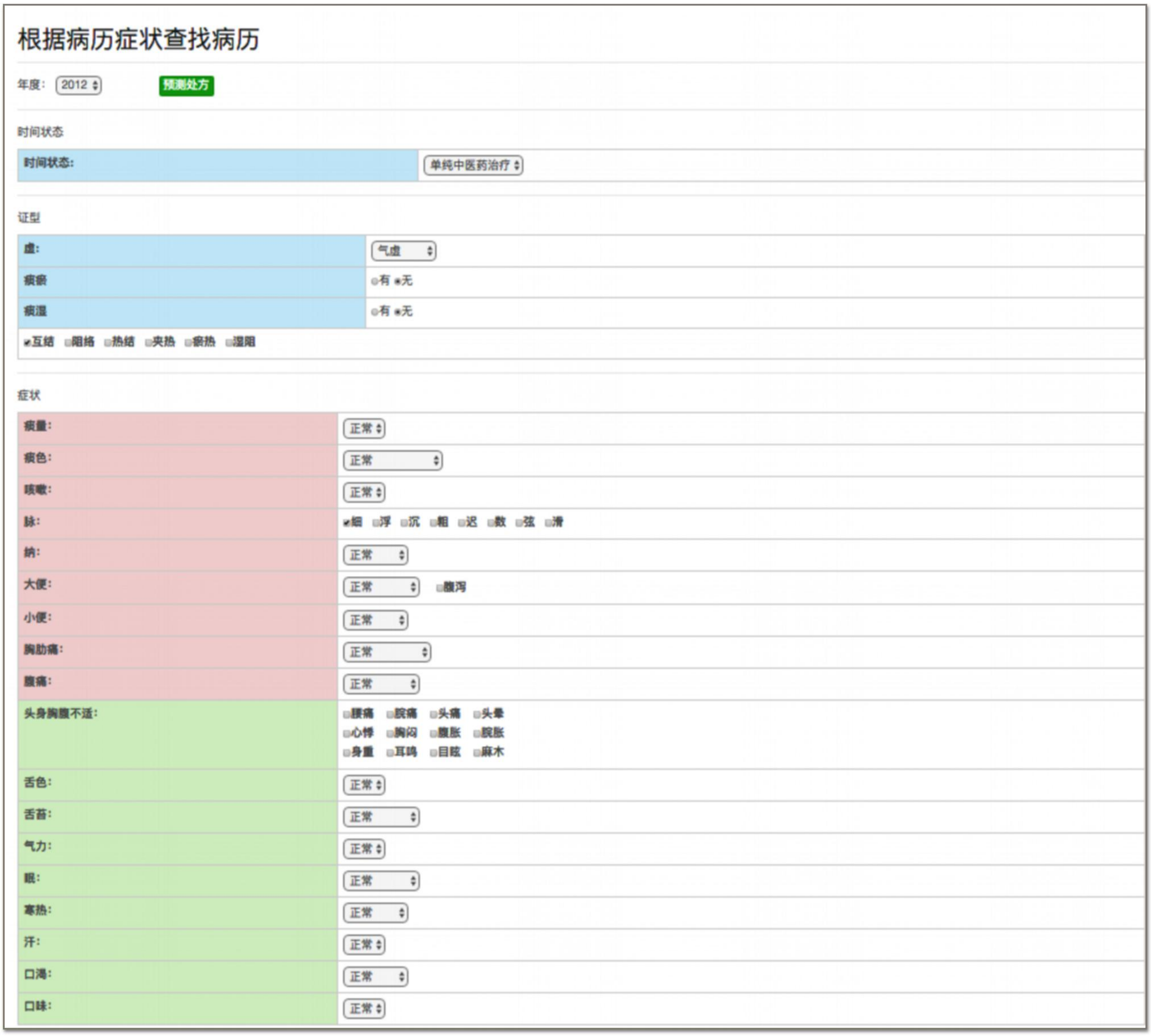
根据病人姓名，处理和中药名称查找病历

年度： 2012 姓名： 处理： 中药名称： 甘草 黄芪 三七片

total count: 3

No.	Info	Detail
1	症状： 气力特差,脉细,红血痰（多）,舌质红色,舌苔黄,咳嗽中,脉弦, 中药处方： 太子参,黄芪,石见穿,红豆杉,白花蛇舌草,藕节炭,甘草,白术,龙葵,茅莓根,薏苡仁,紫珠草,三七片,	Detail
2	症状： 舌质红色,尿频,舌苔黄,脉滑,胸闷,痰白,咳嗽中,脉弦, 中药处方： 党参,黄芪,石见穿,白花蛇舌草,甘草,龙葵,茅莓根,薏苡仁,红豆杉,五味子,三七片,白茅根,紫珠草,	Detail
3	症状： 气力特差,脉细,红血痰（多）,咳嗽重,痰量少,咳嗽中, 中药处方： 太子参,黄芪,石见穿,白花蛇舌草,红豆杉,甘草,白术,龙葵,茅莓根,薏苡仁,三七片,白茅根,紫珠草,	Detail

Figure 3. main page of query result



根据病历症状查找病历

年度： 2012 预测处方

时间状态

时间状态:	单纯中医药治疗

证型

虚:	气虚
痰瘀	○有 ●无
痰湿	○有 ●无
☑互结 □阻络 □热结 □夹热 □瘀热 □湿阻	

症状

痰量:	正常
痰色:	正常
咳嗽:	正常
脉:	☑细 □浮 □沉 □粗 □迟 □数 □弦 □滑
纳:	正常
大便:	正常 □腹泻
小便:	正常
胸胁痛:	正常
腹痛:	正常
头身胸腹不适:	□腰痛 □脘痛 □头痛 □头晕 □心悸 □胸闷 □腹胀 □脘胀 □身重 □耳鸣 □目眩 □麻木
舌色:	正常
舌苔:	正常
气力:	正常
眠:	正常
寒热:	正常
汗:	正常
口渴:	正常
口味:	正常

Figure 4. main page of getting records based on the description conditions

Tool 5. Data Statistics Tools

1. **Statistics the frequency and percent of traditional Chinese medicines in all records. (Figure 5)**

Input:	Batch
Output:	The frequency and percent of traditional Chinese medicines
Function:	Statistics the frequency and percent of traditional Chinese medicines in all records.
API:	method: statisticsMedicinesByBatch() input: batch ——— the batches(or year) of records output: medicines total number percent
Code File:	Java project: EHR Code file: 1. statisticsByCM.jsp 2. StatisticsAction.java

There are a lot of traditional Chinese medicines in all records of each year and many medicines’ percent of frequency are larger than 90%, which means those medicines would exist in almost records. So those medicines can be regarded as the basic medicines of prediction result.

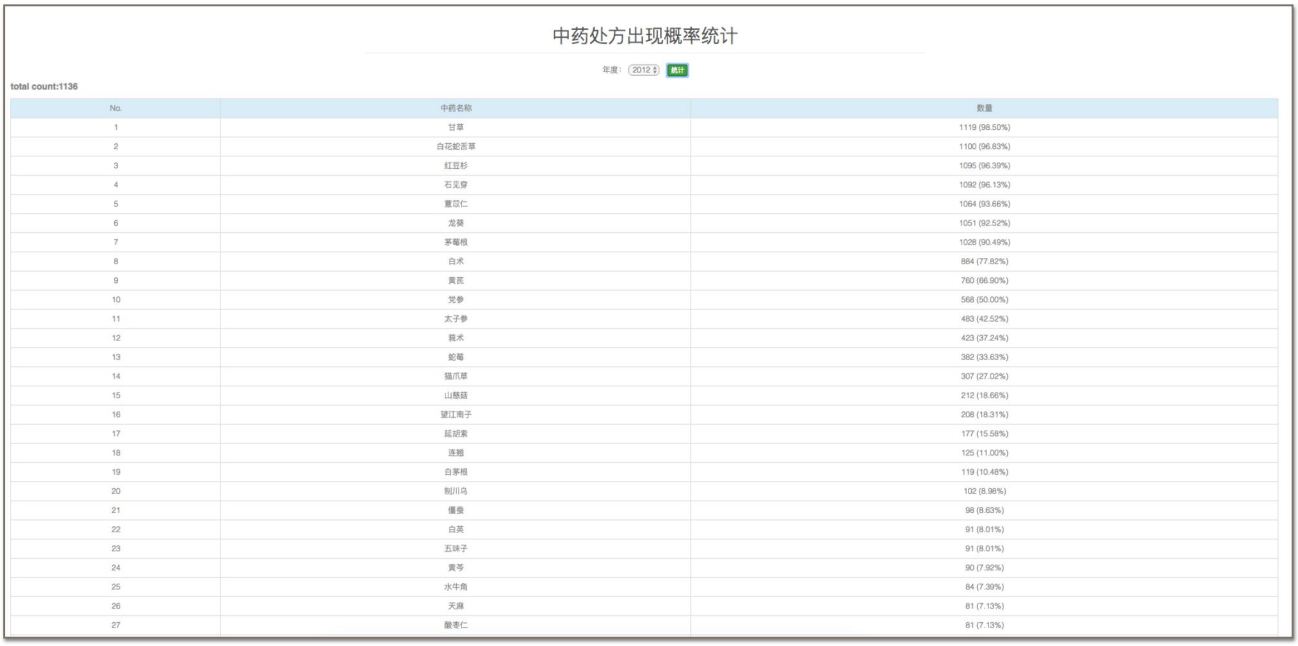
2. Statistics the frequency and percent of traditional Chinese medicine diagnoses in all records. (Figure 6)

Input:	Batch
Output:	The frequency and percent of Chinese medicine diagnoses
Function:	Statistics the frequency and percent of Chinese medicine diagnoses in all records.
API:	method: statisticsByCNDiagnose() input: batch ——— the batches(or year) of records output: chinese medicine diagnoses total number
Code File:	Java project: EHR Code file: 1. statisticsByCNDiagnose.jsp 2. StatisticsAction.java

3. Statistics the frequency and percent of the different combination of traditional Chinese medicines in all records. (Figure 7)

Input:	Batch and Chinese medicines
Output:	The frequency and percent of the different combination of traditional Chinese medicines
Function:	Statistics the frequency and percent of the different combination of traditional Chinese medicines in all records.
API:	method: cnmedicineProba() input: batch ——— the batches(or year) of records medicines ——— the combinations of traditional Chinese medicines output: traditional Chinese medicine combinations total number percent
Code File:	Java project: EHR Code file: 1. cnmedicproba.jsp 2. StatisticsAction.java

This tool shows the statistics results of combinations of traditional Chinese medicines, including medicines combinations (intersection or union), total number and percent. From result of statistics, we can get a lot of rules to predict medicines with rule-based method



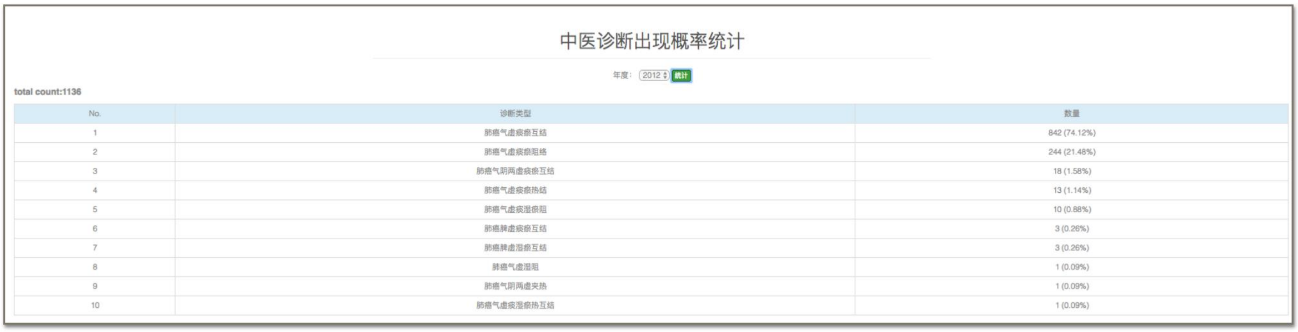
中药处方出现概率统计

年度： 2012

total count:1136

No.	中药名称	数量
1	甘草	1119 (98.50%)
2	白花蛇舌草	1100 (96.83%)
3	红豆杉	1095 (96.39%)
4	石见穿	1092 (96.13%)
5	薏苡仁	1064 (93.66%)
6	龙葵	1051 (92.52%)
7	茅莓根	1028 (90.49%)
8	白术	884 (77.82%)
9	黄芪	760 (66.90%)
10	党参	568 (50.00%)
11	太子参	483 (42.52%)
12	莪术	423 (37.24%)
13	蛇莓	382 (33.63%)
14	猫爪草	307 (27.02%)
15	山慈菇	212 (18.66%)
16	望江南子	208 (18.31%)
17	延胡索	177 (15.58%)
18	连翘	125 (11.00%)
19	白茅根	119 (10.48%)
20	制川乌	102 (8.98%)
21	僵蚕	98 (8.63%)
22	白英	91 (8.01%)
23	五味子	91 (8.01%)
24	黄芩	90 (7.92%)
25	水牛角	84 (7.39%)
26	天麻	81 (7.13%)
27	酸枣仁	81 (7.13%)

Figure 5. main page of statistics tools(1)



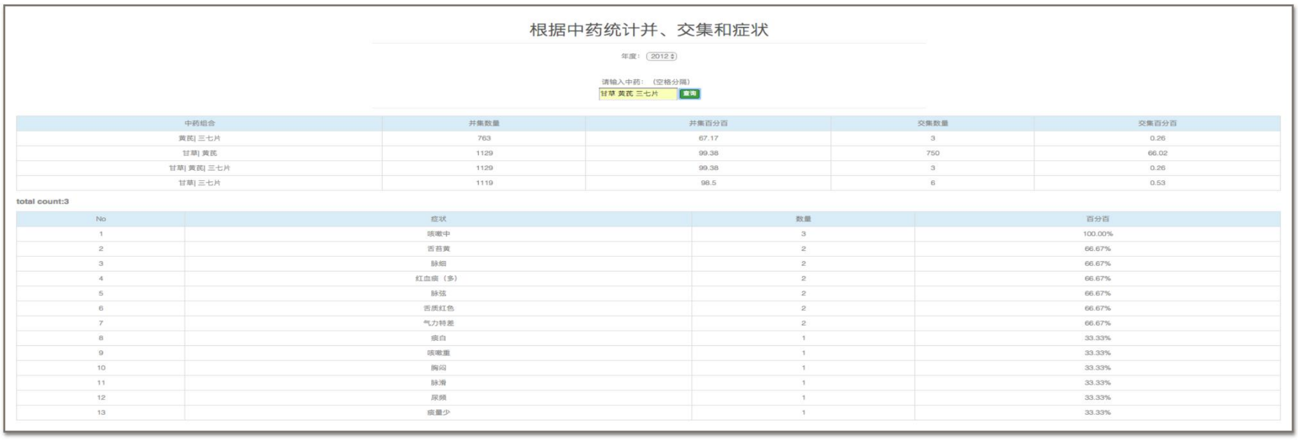
中医诊断出现概率统计

年度： 2012

total count:1136

No.	诊断类型	数量
1	肺癌气虚痰瘀互结	842 (74.12%)
2	肺癌气虚痰瘀阻络	244 (21.48%)
3	肺癌气阴两虚痰瘀互结	18 (1.58%)
4	肺癌气虚痰瘀热结	13 (1.14%)
5	肺癌气虚痰湿瘀阻	10 (0.88%)
6	肺癌脾虚痰瘀互结	3 (0.26%)
7	肺癌脾虚湿瘀互结	3 (0.26%)
8	肺癌气虚湿阻	1 (0.09%)
9	肺癌气阴两虚夹热	1 (0.09%)
10	肺癌气虚痰湿瘀热互结	1 (0.09%)

Figure 6. main page of statistics tools (2)



根据中药统计并、交集和症状

年度： 2012

请输入中药： （空格分隔）

甘草 黄芪 三七片

中药组合	并集数量	并集百分百	交集数量	交集百分百
黄芪| 三七片	763	67.17	3	0.26
甘草| 黄芪	1129	99.38	750	66.02
甘草| 黄芪| 三七片	1129	99.38	3	0.26
甘草| 三七片	1119	98.5	6	0.53

total count:3

No	症状	数量	百分百
1	咳嗽中	3	100.00%
2	舌苔黄	2	66.67%
3	脉细	2	66.67%
4	红血痰（多）	2	66.67%
5	脉弦	2	66.67%
6	舌质红色	2	66.67%
7	气力特差	2	66.67%
8	痰白	1	33.33%
9	咳嗽重	1	33.33%
10	胸闷	1	33.33%
11	脉滑	1	33.33%
12	尿频	1	33.33%
13	痰量少	1	33.33%

Figure 7. main page of statistics tools(3)

Tool 6. Machine Learning data process tool

Input:	All records data stored in MonogDB
Output:	The training data set of Machine Learning with a standard format
Function:	Generate the training data set of Machine Learning with a standard format.
API:	method: cnmedicineProba() input: batch ——— the batches(or year) of records medicines ——— the combinations of traditional Chinese medicines output: traditional Chinese medicine combinations total number percent
Code File:	Java project: CWRelationMapping Code file: 1. com.um.main.PreprocessData.java
Detail:	The training set was divided into two parts: the input part with syndromes and symptoms, and the output part with traditional Chinese medicines. The input part has 53 items, including all syndromes and symptoms in all records. The output part is all traditional Chinese medicines sorted by count in all records. The number of output part is 161.

Tool 7. Traditional Chinese Medicines prediction tool

1. Input-based prediction tool (Figure 8)

Input:	The condition descriptions
Output:	Three prediction methods result
Function:	Use three different methods to predict traditional Chinese medicines: 1. case-based statistical method; 2. machine learning method; 3. rule-based method;
API:	method: predictByInput() input: batch ——— the batches(or year) of records condition description ——— symptoms output: four prediction result of traditional Chinese medicines 1. Case-based statistics prediction result; 2. Machine Learning prediction result; 3. Rule-based prediction result; 4. Comprehensive prediction result based on above results;
Code File:	Java project: EHR Code file: 1. predictMedicine.jsp 2. PredictAciton.java

The number of all items as input symptoms is seventeen and we choice eight items as the main symptoms. They are sleeping, eat, sputum with blood, pain in chest, energy, cough, defecate and diarrhea. Those main symptoms would determine medicines provided by the doctor.

Before predicting medicines by case-based statistics method, we should find similar records with same symptoms or almost same symptoms. At first we choice those records with same symptoms. But if the results are null, we will degrade the limitation of symptoms until we find records with a lot of same symptoms.

The main implementation of input-based prediction tool:

1. Get the syndromes and symptoms from user input;

2. Statistics the occurrence frequency and percent of Chinese medicines in all cases. If the percent of medicine is more than 90%, then it would be one of predicting medicines result;

3. Grouping the remaining Chinese medicines of step 2, and statistics the occurrence percent of each medicine group. If the percent is more than 90%, at least one of this medicine group would exist in the prediction result at least.

4. Get the similar records based on the input syndromes and symptoms, and then statistics all medicines in those similar records. Sort the statistics result by the occurrence percent and remove medicines existed in step 2 result.

5. Choice the medicines existed in result of step 4 util the prediction result is enough to output.

6. Format the input syndromes and symptoms with the standard input structure of Machine Learning, and then get prediction results with Machine Learning algorithm.

7. Format the input syndromes and symptoms with the rule-base structure, and use rule-base part to get the prediction result.

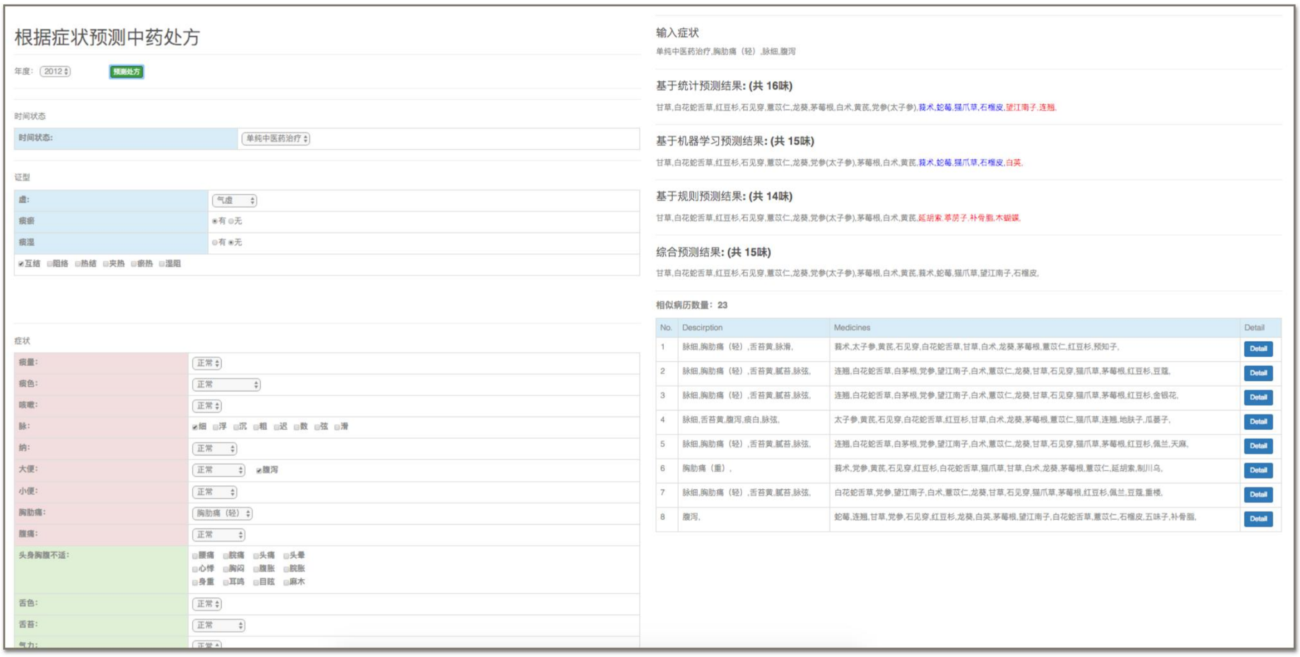
8. Compared with those prediction results of three methods, and choice medicines existed in three results as the comprehensive prediction result, and then choice those medicines existed in two results until the comprehensive result is enough to output.

2. Case-based prediction tool (Figure 9)

Input:	Registration number or order number of records
Output:	The machine learning prediction result.
Function:	Select the exist records as the input of machine learning to predict traditional Chinese medicines.
API:	method: predictByCase() input: count ——— registration number or order number of records output: 1. the traditional Chinese medicines of target record; 2. the prediction result of Machine Learning;
Code File:	Java project: EHR

	Code file: 1. casePredictMedicine.jsp

This tool is used to verify the accuracy of prediction result of machine learning. As result of this tool, we output the origin traditional Chinese medicines in target record, the prediction result of machine learning, the accuracy of prediction result and the detail information of target record.



根据症状预测中药处方

年度: 2012 预测处方

时间状态

时间状态:	单纯中医药治疗

证型

虚:	气虚
痰瘀	● 有 ○ 无
痰湿	○ 有 ● 无
☑ 互结 ☐ 阻络 ☐ 热结 ☐ 夹热 ☐ 瘀热 ☐ 湿阻	

症状

痰量:	正常
痰色:	正常
咳嗽:	正常
脉:	☑ 细 ☐ 浮 ☐ 沉 ☐ 粗 ☐ 迟 ☐ 数 ☐ 弦 ☐ 滑
纳:	正常
大便:	正常 ☑ 腹泻
小便:	正常
胸胁痛:	胸胁痛（轻）
腹痛:	正常
头身胸腹不适:	☐ 腰痛 ☐ 脘痛 ☐ 头痛 ☐ 头晕 ☐ 心悸 ☐ 胸闷 ☐ 腹胀 ☐ 脘胀 ☐ 身重 ☐ 耳鸣 ☐ 目眩 ☐ 麻木
舌色:	正常
舌苔:	正常
气力:	正常

输入症状

单纯中医药治疗,胸胁痛（轻）,脉细,腹泻

基于统计预测结果: (共 16味)

甘草,白花蛇舌草,红豆杉,石见穿,薏苡仁,龙葵,茅莓根,白术,黄芪,党参(太子参),莪术,蛇莓,猫爪草,石榴皮,望江南子,连翘,

基于机器学习预测结果: (共 15味)

甘草,白花蛇舌草,红豆杉,石见穿,薏苡仁,龙葵,党参(太子参),茅莓根,白术,黄芪,莪术,蛇莓,猫爪草,石榴皮,白英,

基于规则预测结果: (共 14味)

甘草,白花蛇舌草,红豆杉,石见穿,薏苡仁,龙葵,党参(太子参),茅莓根,白术,黄芪,延胡索,葶苈子,补骨脂,木蝴蝶,

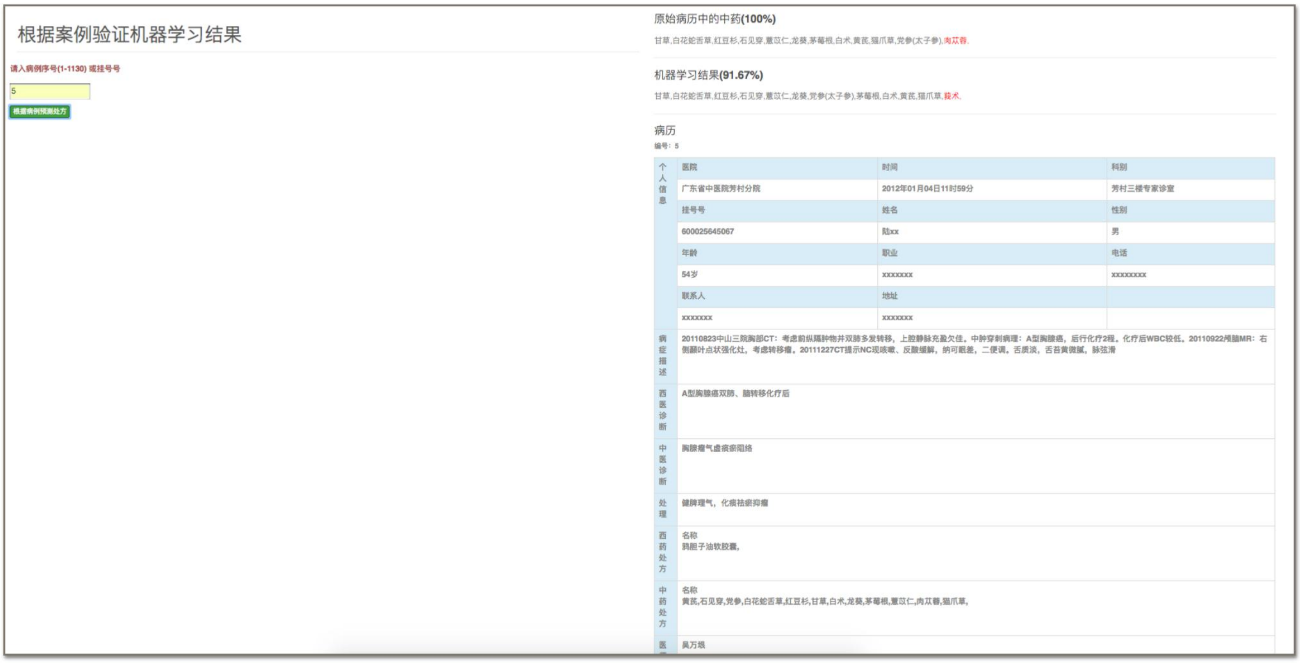
综合预测结果: (共 15味)

甘草,白花蛇舌草,红豆杉,石见穿,薏苡仁,龙葵,党参(太子参),茅莓根,白术,黄芪,莪术,蛇莓,猫爪草,望江南子,石榴皮,

相似病历数量: 23

No.	Descirption	Medicines	Detail
1	脉细,胸胁痛（轻）,舌苔黄,脉滑,	莪术,太子参,黄芪,石见穿,白花蛇舌草,甘草,白术,龙葵,茅莓根,薏苡仁,红豆杉,预知子,	Detail
2	脉细,胸胁痛（轻）,舌苔黄,腻苔,脉弦,	连翘,白花蛇舌草,白茅根,党参,望江南子,白术,薏苡仁,龙葵,甘草,石见穿,猫爪草,茅莓根,红豆杉,豆蔻,	Detail
3	脉细,胸胁痛（轻）,舌苔黄,腻苔,脉弦,	连翘,白花蛇舌草,白茅根,党参,望江南子,白术,薏苡仁,龙葵,甘草,石见穿,猫爪草,茅莓根,红豆杉,金银花,	Detail
4	脉细,舌苔黄,腹泻,痰白,脉弦,	太子参,黄芪,石见穿,白花蛇舌草,红豆杉,甘草,白术,龙葵,茅莓根,薏苡仁,猫爪草,连翘,地肤子,瓜蒌子,	Detail
5	脉细,胸胁痛（轻）,舌苔黄,腻苔,脉弦,	连翘,白花蛇舌草,白茅根,党参,望江南子,白术,薏苡仁,龙葵,甘草,石见穿,猫爪草,茅莓根,红豆杉,佩兰,天麻,	Detail
6	胸胁痛（重）,	莪术,党参,黄芪,石见穿,红豆杉,白花蛇舌草,猫爪草,甘草,白术,龙葵,茅莓根,薏苡仁,延胡索,制川乌,	Detail
7	脉细,胸胁痛（轻）,舌苔黄,腻苔,脉弦,	白花蛇舌草,党参,望江南子,白术,薏苡仁,龙葵,甘草,石见穿,猫爪草,茅莓根,红豆杉,佩兰,豆蔻,重楼,	Detail
8	腹泻,	蛇莓,连翘,甘草,党参,石见穿,红豆杉,龙葵,白英,茅莓根,望江南子,白花蛇舌草,薏苡仁,石榴皮,五味子,补骨脂,	Detail

Figure 8. prediction tool main page



根据案例验证机器学习结果

请入病例序号(1-1130) 或挂号号

5

根据病例预测处方

原始病历中的中药(100%)

甘草,白花蛇舌草,红豆杉,石见穿,薏苡仁,龙葵,茅莓根,白术,黄芪,猫爪草,党参(太子参),肉苁蓉,

机器学习结果(91.67%)

甘草,白花蛇舌草,红豆杉,石见穿,薏苡仁,龙葵,党参(太子参),茅莓根,白术,黄芪,猫爪草,莪术,

病历

编号: 5

个人信息	医院	时间	科别
	广东省中医院芳村分院	2012年01月04日11时59分	芳村三楼专家诊室
	挂号号	姓名	性别
	600025645067	陆xx	男
	年龄	职业	电话
	54岁	xxxxxxx	xxxxxxxx
	联系人	地址	
	xxxxxxx	xxxxxxx	
病症描述	20110823中山三院胸部CT：考虑前纵隔肿物并双肺多发转移，上腔静脉充盈欠佳。中肿穿刺病理：A型胸腺癌，后行化疗2程。化疗后WBC较低。20110922颅脑MR：右侧颞叶点状强化灶，考虑转移瘤。20111227CT提示NC现咳嗽、反酸缓解，纳可眠差，二便调。舌质淡，舌苔黄微腻，脉弦滑		
西医诊断	A型胸腺癌双肺、脑转移化疗后		
中医诊断	胸腺瘤气虚痰瘀阻络		
处理	健脾理气，化痰祛瘀抑瘤		
西药处方	名称 鸦胆子油软胶囊,		
中药处方	名称 黄芪,石见穿,党参,白花蛇舌草,红豆杉,甘草,白术,龙葵,茅莓根,薏苡仁,肉苁蓉,猫爪草,		
医	吴万垠		

Figure 9. main page of machine learning prediction tool based on existed case

Appendix 1:

1. The structure of XML

```
<ehealthrecord>
        <hospital>hospital</hospital>
        <medicineservice>department</medicineservice>
        <date>date</date>
        <registrationno>registration</registrationno>
        <doctor>doctor</doctor>
        <patientinfo>
        <!--patient basic information-->
            <name>name</name>
            <age>age</age>
            <gender>gender</gender>
            <profession>profession</profession>
            <phone>phonenumber</phone>
            <contact>contact</contact>
            <address>address</address>
        </patientinfo>
        <conditionsdescribed>conditiondescribed</conditionsdescribed>
        <diagnostics>
        <!--diagnose-->
            <westerndiagnostics>western medicine diagnose</westerndiagnostics>
            <chinesediagnostics>chinese medicine diagnose</chinesediagnostics>
        </diagnostics>
        <process>process methods</process>
        <medicine>
        <!--medicines-->
            <westernMedicines>
                <westernMedicine>
                    <group>group</group>
                    <wname>name</wname>
                    <specifications>specification</specifications>
                    <usage>usage</usage>
                    <amount>amount</amount>
                </westernMedicine>
            </westernMedicines>
            <chineseMedicines>chinese medicines
                <chineseMedicine>
                    <cname>name</cname>
                    <bias1>bias1</bias1>
                    .....
                    <number>number</number>
                    <unit>unit</unit>
                </chineseMedicine>
            </chineseMedicines>
        </medicine>
    </ehealthrecord>
```