

Orvosi szövegértelmező rendszer

Témavezető:

dr. ing. Johann Stan

Hallgató:

Aszalos Zoltán

Célkitűzések

- Gyógyszerek közötti kölcsönhatások kimutatása
- Nagy mennyiségű gyógyszer adathalmaz feldolgozása
- Orvosi kifejezések meghatározása
- Különböző szerepű rendszerek integrálása
- Webalkalmazás megvalósítása

Bevezetés

- Gyógyszerkölsönhatások szerepe
- Adatforrás: a DailyMed portál által publikált XML dokumentumok

```
<document>
  <manufacturedProduct>
    <code code="50066-374" codeSystem="2.16.840.1.113883.6.69"/>
    <name>Dragon<suffix>Tabs</suffix></name>
    ...
    <ingredientSubstance>
      <code code="WK2XYI10QM" codeSystem="2.16.840.1.113883.4.9"/>
      <name>IBUPROFEN</name>
    </ingredientSubstance>
    ...
  <component>
    <section ID="ID_71cd2e00-0c94-4b47-8f16-96ef2f99fb53">
      <id root="71cd2e00-0c94-4b47-8f16-96ef2f99fb53"/>
      <code code="34071-1" codeSystem="2.16.840.1.113883.6.1" displayName="WARNINGS SECTION"/>
      <title>Warnings</title>
      <text>
        <paragraph>
          <content styleCode="bold">Allergy alert</content>: Ibuprofen may cause a severe allergic reaction,
          especially in people allergic to aspirin. Symptoms may include: ■ hives ■ facial swelling ■ asthma (wheezing) ■
          shock<content styleCode="bold"/>■ skin reddening <br/>■ rash ■ blisters<br/>If an allergic reaction occurs, stop use
          and seek medical help right away.<br/>
          ...
        </paragraph>
      </text>
    </section>
  </component>
</document>
```

Eszközök

- UMLS (Unified Medical Language System)
- Az UMLS komponensei:
 - Metathesaurus
 - Semantic Network
 - SPECIALIST Lexicon
- MetaMap: eszköz UMLS kifejezések felismerésére
- Hadoop Pig
- SPARQL

MetaMap

- Futtatás: terminálból vagy Java API által
- Példa:

echo "Ibuprofen may cause a severe allergic reaction, especially in people allergic to aspirin." | ./metamap14 -J sosy,dsyn,orch,phsu,inpo --XMLf > out.xml

sosy	Sign or Symptom
dsyn	Disease or Syndrome
orch	Organic Chemical
phsu	Pharmacologic Substance
inpo	Injury or Poisoning

MetaMap eredménye

...

<Mappings Count="1">

<Mapping>

<MappingScore>-623</MappingScore>

<MappingCandidates Total="1">

<Candidate>

<CandidateScore>-623</CandidateScore>

<CandidateCUI>C0004057</CandidateCUI>

<CandidateMatched>Aspirin</CandidateMatched>

<CandidatePreferred>Aspirin</CandidatePreferred>

<MatchedWords Count="1">

<MatchedWord>aspirin</MatchedWord>

</MatchedWords>

<SemTypes Count="2">

<SemType>orch</SemType>

<SemType>phsu</SemType>

...

Hadoop

- Keretrendszer nagy mennyiségű adat (big data) kezelésére és feldolgozására
- Komponensei: Common, HDFS, YARN, MapReduce
- Pig: eszköz MapReduce feladatok (job) definiálására
- Pig Latin: szkriptnyelv, adatforrása a HDFS, adatfolyamokat határoz meg

Pig szkript példa

```
drugs = LOAD 'mongodb://localhost:27017/drugsdb.drugs'
        USING com.mongodb.hadoop.pig.MongoLoader('name: chararray,
producerId: chararray, producerName: chararray, ingredients');

drugs2 = FOREACH drugs GENERATE name, producerId, producerName,
js.createSubstancesBag(ingredients) AS ingredients:{t:(id, name, code)};

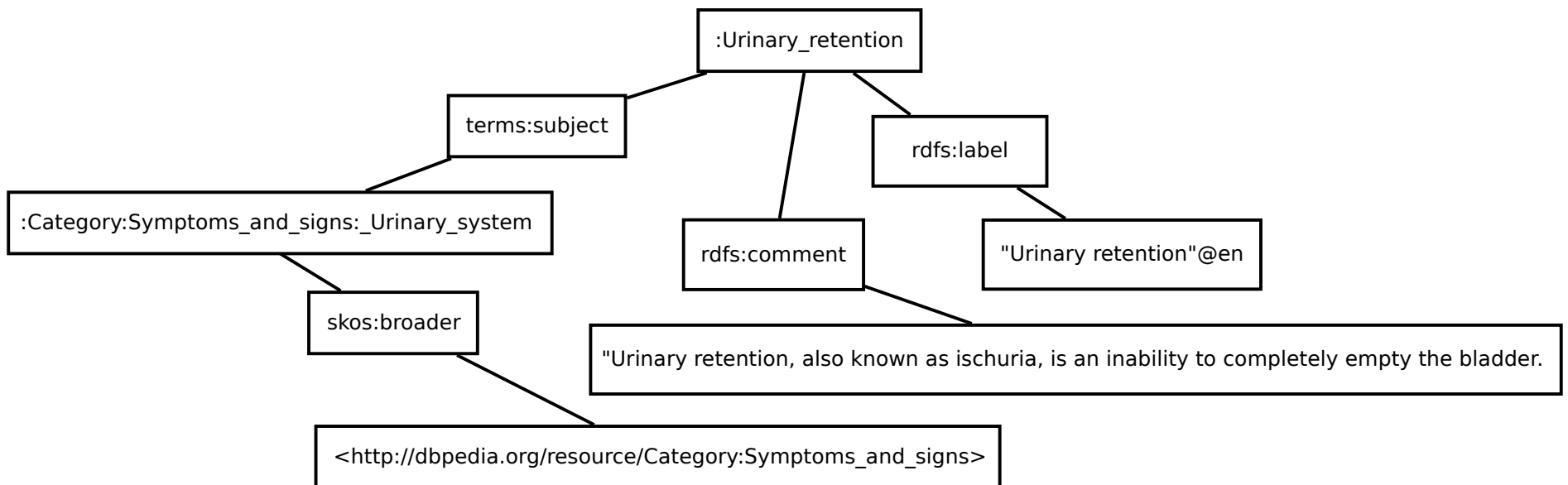
producersGrouped = GROUP drugs2 BY producerId;

producersDrugs = FOREACH producersGrouped GENERATE
    MIN(drugs2.producerId) AS producerId,
    MIN(drugs2.producerName) AS producerName,
    drugs2.(name, ingredients) AS drugs:{t:(name, ingredients)};

STORE producersDrugs INTO 'mongodb://localhost:27017/drugsdb.producers'
USING com.mongodb.hadoop.pig.MongoStorage();
```


SPARQL

- SPARQL Protocol and RDF Query Language
- Részbe a szemantikus webnek
- SPARQL csomópont <http://dbpedia.org/snorql/>



SPARQL példa

```
PREFIX dbo: <http://dbpedia.org/ontology/>
```

```
PREFIX terms: <http://purl.org/dc/terms/>
```

```
SELECT DISTINCT *
```

```
WHERE {" +
```

```
{
```

```
    ?s skos:broader
```

```
<http://dbpedia.org/resource/Category:Symptoms_and_signs> .
```

```
    ?s2 terms:subject ?s .
```

```
    ?s2 rdfs:label ?label .
```

```
    ?s2 rdfs:comment ?comment .
```

```
    FILTER (regex(?label, 'KEYWORD', 'I'))
```

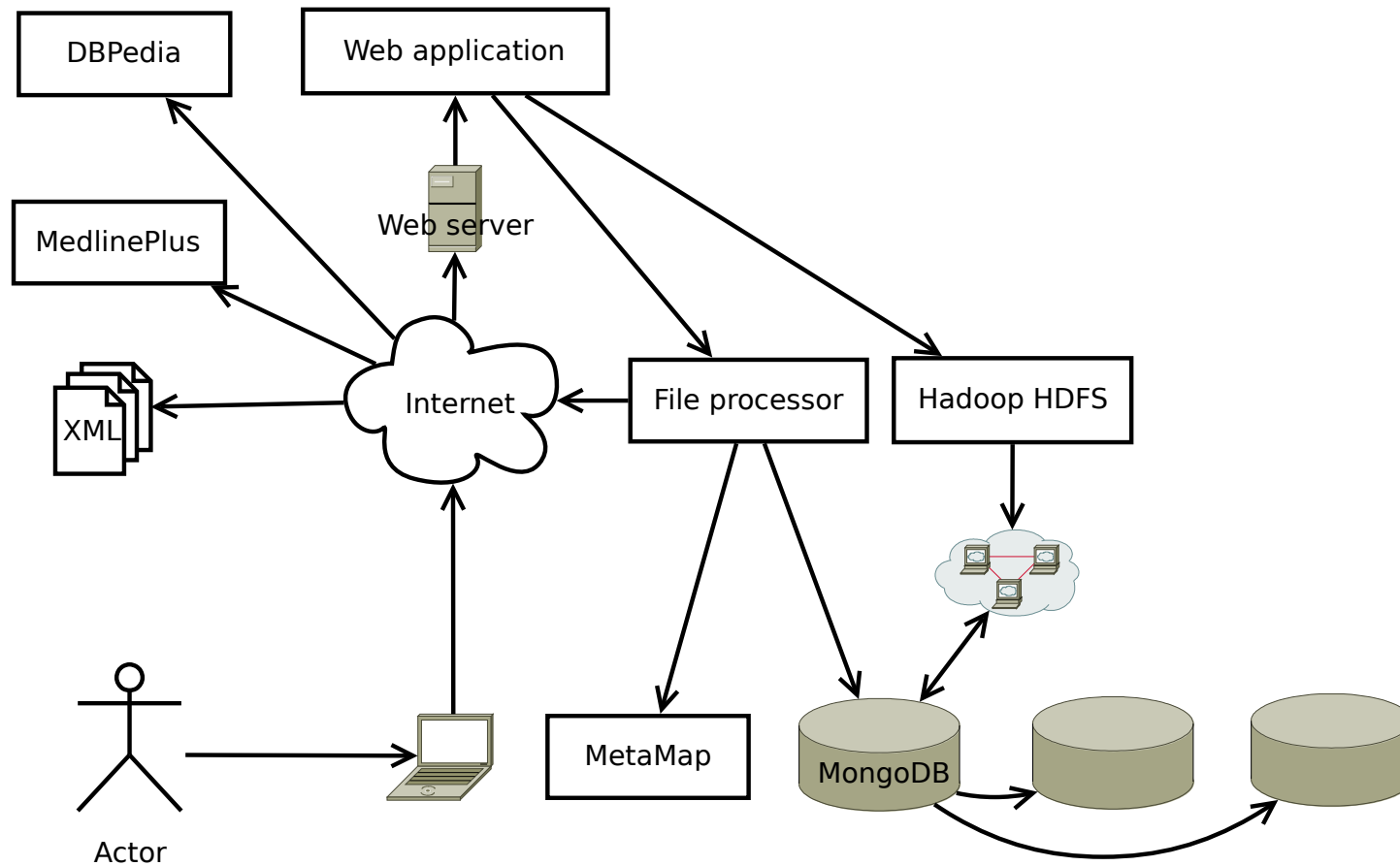
```
    FILTER (lang(?label) = 'en')
```

```
    FILTER (lang(?comment) = 'en')
```

```
}
```

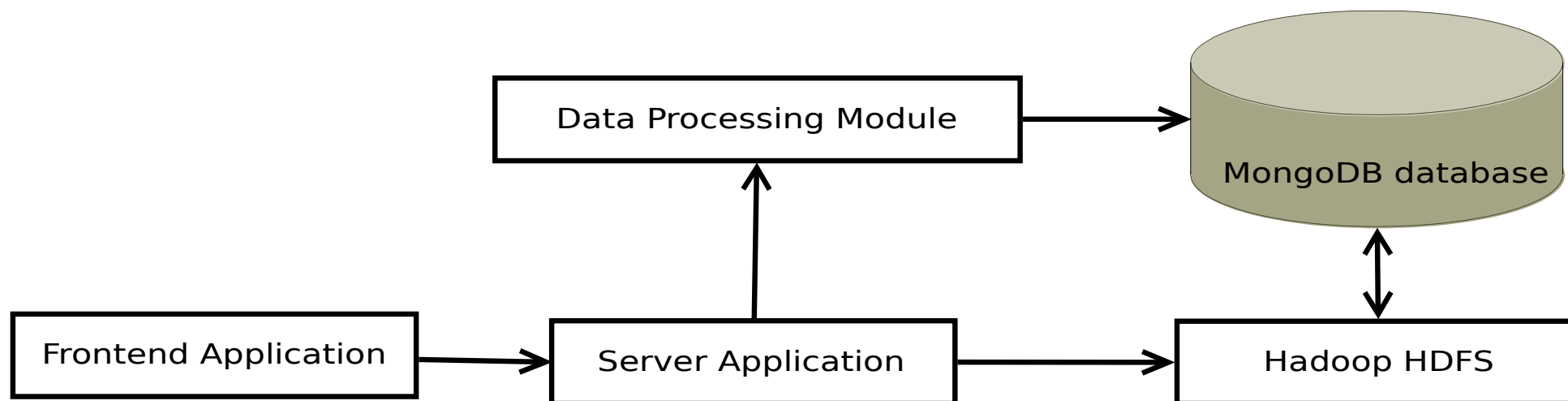
```
...
```

A rendszer áttekintése

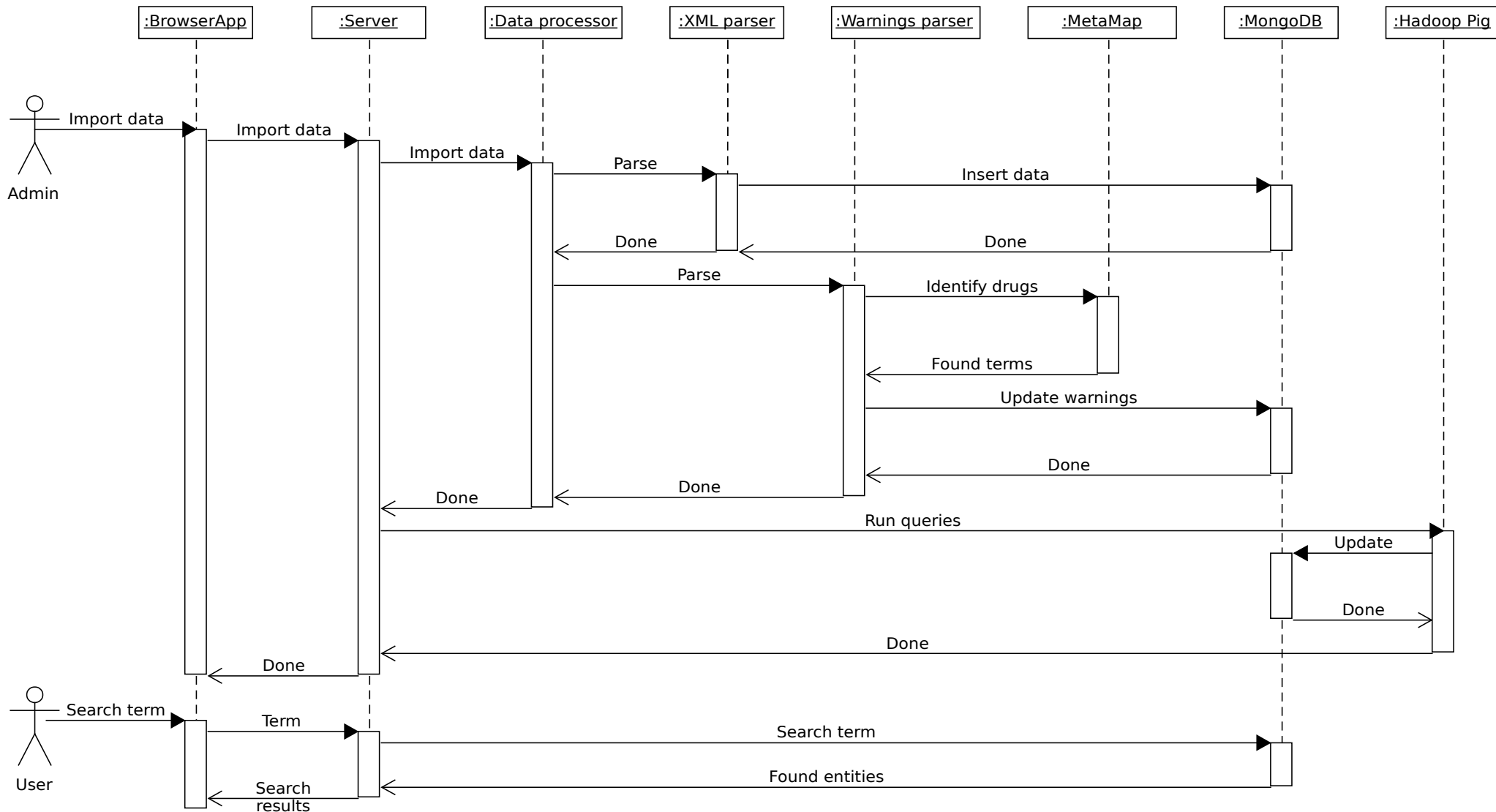


A rendszer architektúrája

- Adatfeldolgozó komponens
- Szerver komponens
- Hadoop HDFS
- MongoDB adatbázis
- Kliens komponens



A rendszer szekvenciadiagramja



Adatfeldolgozó komponens funkciói

- XML dokumentumok beolvasása, kitömörítése
- adatok elemzése: gyógyszer, gyártó, alapanyagok, útmutatók, figyelmeztetések
- adatok beszúrása az adatbázisba
- MetaMap eszköz futtatása és eredmények elemzése
- adatbázis frissítése a találatokkal

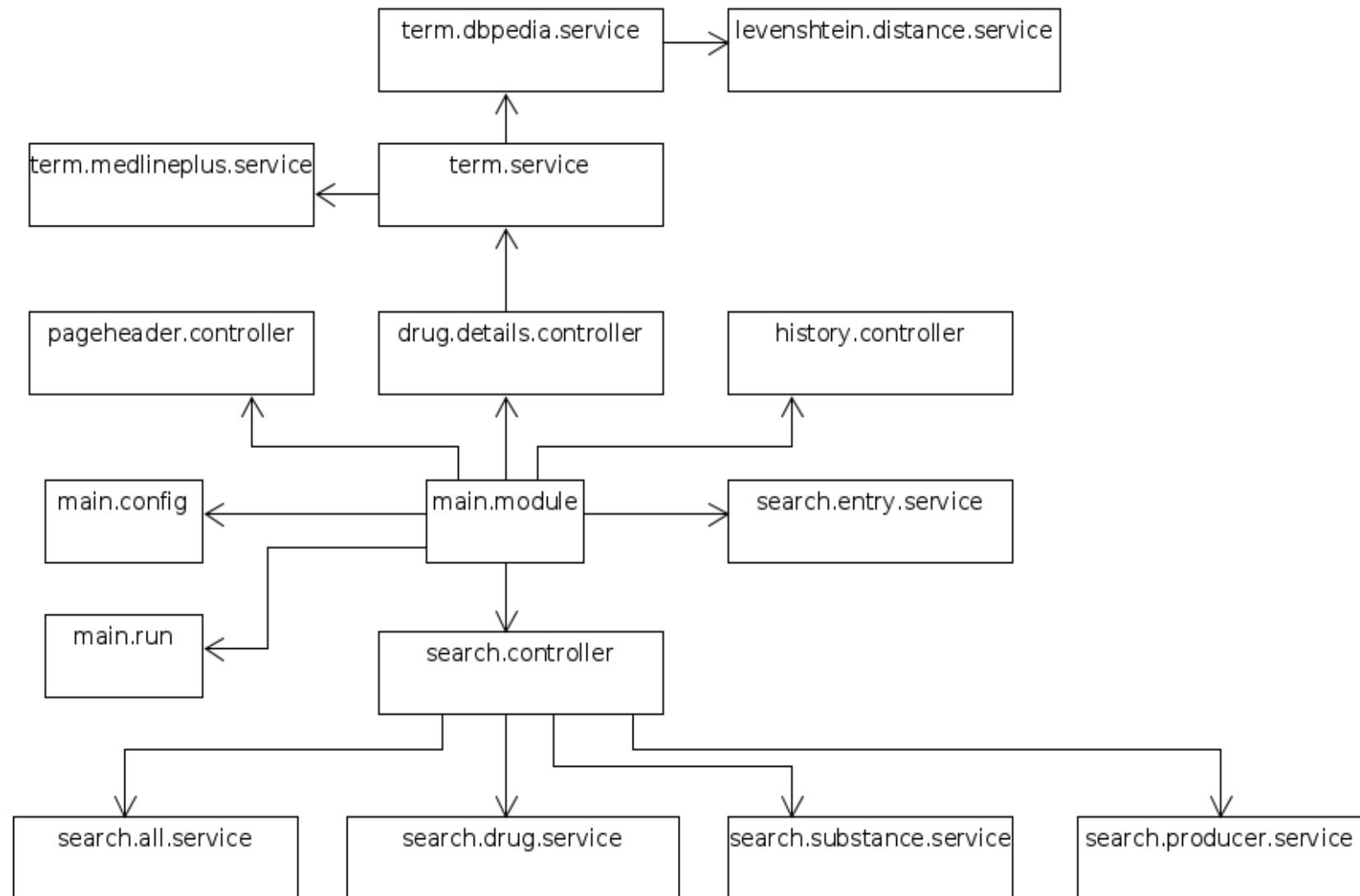
Szerver komponens funkciói

- Elérhetővé teszi az alkalmazást az Interneten
- REST végpontok fenntartása: user, main, search entry
- adatfeldolgozó komponens aktiválása
- Pig lekérdezések futtatása

Kliens (frontend) komponens

- Böngészőben futó alkalmazást valósítja meg
- SPA
- Moduláris architektúrát követ: admin, common, main, users, widgets
- Funkciói:
 - felhasználó regisztrálása
 - keresések biztosítása: gyógyszer, gyártó, alapanyag
 - gyógyszerköölcsönhatások kimutatása
 - előzmények megjelenítése

Main modul road map diagramja



Adatfeldolgozás aktiválása

DrugsDB

Admin

Search

History

Profile

Contact

a

Log out



Import data into the database

Start



Gyógyszer keresése

Search by drug ▾ dragon  

#	Name	Producer	Ingredients 	
1	DragonTabs	Genomma Lab USA	BAPTISIA TINCTORIA ROOT IBUPROFEN SILICON DIOXIDE STARCH, CORN HYPROMELLOSES FERRIC OXIDE RED MAGNESIUM STEARATE CELLULOSE, MICROCRYSTALLINE SODIUM STARCH GLYCOLATE TYPE A CORN TITANIUM DIOXIDE LACTOSE POLYDEXTROSE POLYETHYLENE GLYCOLS	

Gyógyszerkölcshatások megjelenítése

DrugsDBAdminSearchHistoryProfileContact

aLog out

Drug details

Name

DragonTabs

Producer

Genomma Lab USA

Ingredients

BAPTISIA TINCTORIA ROOT; IBUPROFEN; SILICON DIOXIDE; STARCH, CORN; HYPROMELLOSES; FERRIC OXIDE RED; MAGNESIUM STEARATE; CELLULOSE, MICROCRYSTALLINE; SODIUM STARCH GLYCOLATE TYPE A CORN; TITANIUM DIOXIDE; LACTOSE; POLYDEXTROSE; POLYETHYLENE GLYCOLS;

Interactions

[Ibuprofen](#); [Aspirin](#); [Anti-Inflammatory Agents, Non-Steroidal](#); [Nonsteroidal Anti-inflammatory Drug \[EPC\]](#); [Anticoagulants](#); [Anti-coagulant \[EPC\]](#); [Steroid containing root canal medicament](#); [Pharmaceutical Preparations](#); [Drugs, Non-Prescription](#); [Naproxen](#); [Diuretics](#); [Pharmacologic Substance](#);

Additional information

Warnings

[Do not use](#)

[Ask a doctor before use if](#)

[Ask a doctor or pharmacist before use if](#)

[When using this product](#)

[Stop use and ask a doctor if](#)

[If pregnant or breast-feeding](#)

[Do not use](#)

[Ask a doctor before use if](#)

[Ask a doctor or pharmacist before use if](#)

Allergy alert Ibuprofen may cause a severe allergic reaction, especially in people allergic to aspirin. Symptoms may include hives facial swelling asthma (wheezing) shock skin reddening rash blistersIf an allergic reaction occurs, stop use and seek medical help right away. Stomach bleeding warning This product contains a nonsteroidal anti-inflammatory drug (NSAID), which may cause severe stomach bleeding. The chances are higher if you are age 60 or older have had stomach ulcers or bleeding problems take a blood thinning (anticoagulant) or steroid drug take other drugs containing prescription or nonprescription NSAIDs (aspirin, ibuprofen, naproxen, or others) have 3 or more alcoholic drinks every day while using this product take more or for a longer time than directed.

#	Keyword	Preferred keyword	Type	Description
1	Allergy	Allergic disposition	disease or syndrome	
2	Allergic drug	Drug Allergy	disease or	A drug allergy is an allergy to a drug, most commonly a medication. Medical attention should be sought immediately if an allergic reaction is suspected. An allergic reaction will not occur on the first exposure to a substance. The first exposure allows the body to create antibodies.

Eredmények

- Találatok minősítése:
 - true positive (TP)
 - true negative (TN)
 - false positive (FP)
 - false negative (FN)

	Helyes	Helytelen
Kiválasztott gyógyszerek	TP	FP
Kihagyott gyógyszerek	FN	TN

- Példa: „Aspirin has adverse effects if administered in combination with Arava but it is safe to take it with multivitamins.”

Eredmények: kölcsönhatások felismerése

- Adatforrás: 10 gyógyszer számos kölcsönhatással

Típus	Képlet	Érték
Teljes pontosság (accuracy)	$\frac{TP + TN}{TP + TN + FP + FN}$	81%
Szelektív precizitás (precision)	$\frac{TP}{TP + FP}$	82%
Helyesség (recall)	$\frac{TP}{TP + FN}$	96%

Eredmények: betegségek és tünetek felismerése

- Adatforrás: 10 gyógyszer
- Választott mérték:
$$\frac{\text{helyes találatok száma}}{\text{helyes találatok száma} + \text{hiányzó találatok száma}}$$

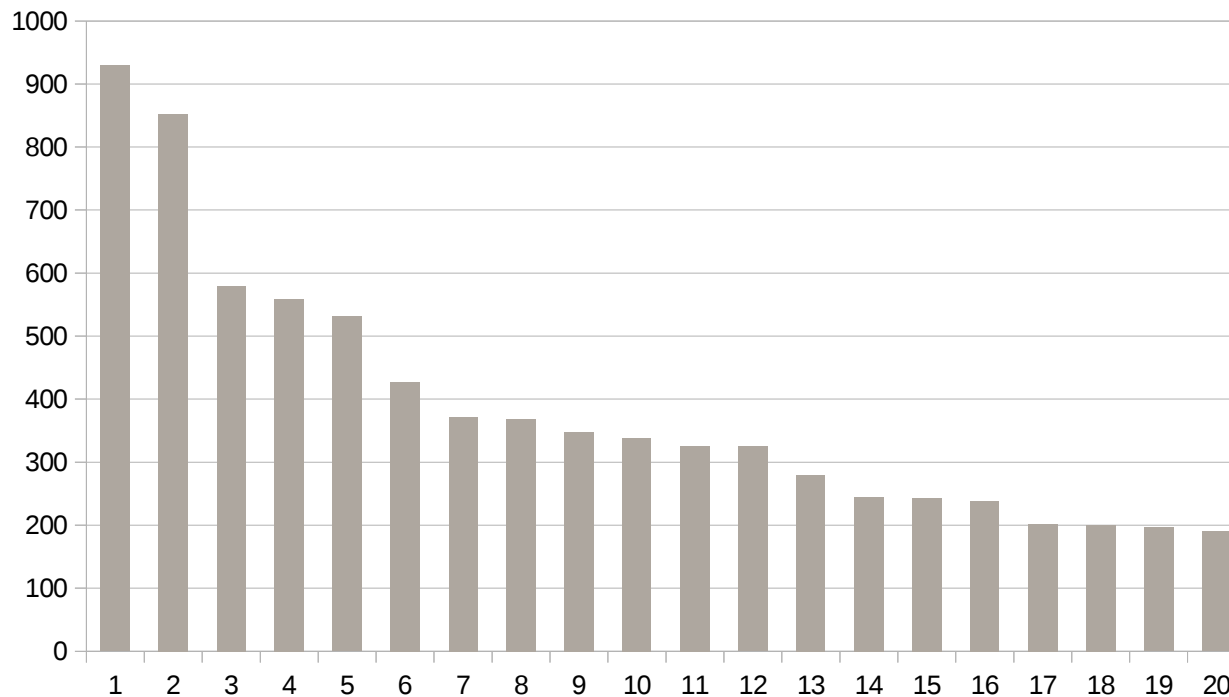
Típus	Precizitás
MedlinePlus	82%
DBPedia	56%

Eredmények: betegségek és tünetek előfordulása

- Adatforrás: a DailyMed portál 2016. januárjában kiadott gyógyszer adathalmaza

Sorszám	Betegség, tünet találat neve (candidate matched)	Betegség, tünet ajánlott neve (candidate preferred)	Típus (semantic type)
1	Ophthalmia	Endophthalmitis	dsyn
2	Rash	Exanthema	dsyn
3	Ill	Malaise	sosy
4	Vomiting	Vomiting	sosy
5	Nausea	Nausea	sosy
6	Swelling	Edema	patf
7	Asthma	Asthma	dsyn
8	Diabetes	Diabetes mellitus	dsyn
9	Diarrhea	Diarrhea	sosy
10	Arrest	Cardiac Arrest	dsyn
11	Dizziness	Dizziness	sosy
12	Headache	Headache	sosy
13	Heart disease	Heart disease	dsyn
14	Liver disease	Liver disease	dsyn
15	Sore Throat	Sore Throat	sosy
16	Emphysema	Pulmonary Emphysema	patf
17	Bronchitis, Chronic	Bronchitis, Chronic	dsyn
18	Bleeding	Hemorrhage	patf
19	Seizures	Seizures	sosy
20	Stroke	Cerebrovascular accident	dsyn

Eredmények: betegségek és tünetek előfordulása



Leggyakrabban előforduló kifejezések:
ophtalmia, rash, ill, vomiting, nausea, swelling,
asthma, stb.

Eredmények: kliens komponens tesztelése

- Felhasznált mérték: kódbázis lefedettségének a vizsgálata egység tesztek által
- Kritériumok: utasítások, elágazások, függvények, sorok

Kifejezések	84.29%	118/140
Elágazások	70.59%	24/34
Függvények	72.73%	32/44
Sorok	84.29%	118/140

Következtetések és tervek

- Technológiák sikeres integrálása
- MetaMap eszköz alkalmazása
- Nyílt forrású eszközök használata
- Javascript nyelv széleskörű támogatása
- MetaMap találatok szűrése, feketelista bevonása
- Orvosi kifejezések keresésének optimalizálása
- Tesztesetek definiálása a teljes kódbázisra
- Offline mód támogatása