Finding Generic Drugs

interlinking DBPedia with FreeDB based on drug brands

written by
Markus Dietsche 1513384
Dandan Li 1486051

submitted to the
Data and Web Science Group
Prof. Dr. Paulheim
University of Mannheim

December 2015

Contents

Chapter 1

Application domain and goals

This documents describes the development of a generics search engine, based on the publicly available knowledge bases DBPedia[1] and FreeBase[2]. Furthermore it discusses the limitations of the corresponding TRL6[3] prototype and contains proposals for further development.

Problem Currently people are forced to rely on expert opinion to find a substitute, e.g a cheaper alternative, to their medication. Since many experts, like doctors or pharmacists, are granted benefits from pharmaceutical companies their judgment is often assorted to be biased. Furthermore drugs often have fantasy names, due to marketing reasons. This is in particular the case for over-the-counter drugs. Therefore a comparison is only possible based ingredients. In case of drugs without patent protection this potentially results in filtering thousands of different drugs by their ingredients. Needless to say that even well informed an unbiased expert will not know all generics by heart.

Solution The search engine is designed for private individuals, without any medical expert knowledge. By entering a drug brand name or active substance the prototype will return a list of identical substitute drugs and their corresponding drug manufacturers.

[1] `www.dbpedia.org`

[2] `www.freebase.com`

[3] Technology Readiness Level `http://resources.sei.cmu.edu/library/asset-view.cfm?assetID=5835`

Chapter 2

Datasets

Datasets used

DBPedia

FreeBase

Access methods

Combination of knowledge bases

Chapter 3

Techniques Used

TODO

Chapter 4

Example results

Outcome

Queries

Chapter 5

Known limitations

Mu

Current manufacturer FreeBase does not consider time dimension. Therefore multiple drug manufacturers can be found for the same brand if the brand was e.g. sold. An example is Ecotrin (-¿ page 4) which was developed and produced by GlaxoSmithKline. In 2011 it was sold in bundle deal to Prestige Brands, which is its current manufacturer. Workaround: crawl information from trademark registers. Add information about production time frames to output, since drugs produced from previous brand owners could still be queried.

Quantity of active substance DBPedia does refer all generics to a single entity, it does not contain information about the quantity of active substances. Therefore a 500mg Aspirin is considered the same as a 200mg Aspirin. Workout: filter quantity from user input and check it on corresponding FreeDB pages.

Impossible queries

Multi ingredient drugs No multi ingredient drugs e.g. ”Neo Citran”. Because DBPedia does not consider it as a SameAs relationship -¿ no generics found.

Drugs not listen as sameAs If a drug is not linked correctly on DBPedia, e.g. it has an own page, even if the ingredients are identical, the search wont function.

Lessons learned

Use datasets with multiple maintainers since our initial project proposal, searching manufacturers of certain product categories and visualizing them on Google maps, had to be discarded, since ProductDB was shut down and the only maintainer was unreachable.

Challenges 1. SameAs 2. Time

Biggest Obstacles

Appendix A

Program Code / Resources

The source code can be obtained via git. Access to the private repository will be granted upon request by silviadali101@gmail.com