Open Self Medication on LOD

Olivier Curé

Université Paris-Est, LIGM, CNRS UMR 8049, France
ocure@univ-mlv.fr

Abstract. Open Self Medication[1] is a Web application that aims to accompany the general public in his initiative to self-medicate, i.e. the act of treating undiagnosed medical ailments with unprescribed drugs. The application achieves this goal by providing a set of functionalities that ensure safety and efficiency of this practice. With safety, we mean that the system guides the end-user from a set of common mild medical signs to adapted molecules and drug products, but also highlights the risks, e.g. drug interactions, adverse events, of self prescribing a drug in a given situation. The efficiency argument corresponds to providing a rating, based on a tolerance/efficiency ratio designed by a team of health care professionals, to some identified self-medication molecules. A main characteristic of this application is that almost all the data processed by the system and presented to the end-user comes from a subset of the LOD datasets, namely *DrugBank*, *DailyMed*, *Sider* and *DBPedia*. This paper motivates the design of such an application, provides the main design choices, describes some implementation details and presents lessons learned and future work.

1 Motivation

In OECD (Organization for Economic Co-operation and Development) countries, due to the emergence and wide distribution of new drugs as well as the aging of populations, the consumption of pharmaceuticals keeps increasing every year and in 2009, the associated bill has been estimated to USD 700 billion, i.e. 19% of the total health spending [6]. It is considered that this rise of drug products consumption has not been followed by an improvement of the general public's medical knowledge. That is, the general public totally relies on the skills of health care professionals when it comes to consuming drugs. This is quite risky since this same OECD document reports that self-medication or over-the-counter (OTC) pharmaceutical products typically account for around 15% of the total spending. This is exacerbated since 2008, the beginning of the financial crisis, with self-medication, i.e. the act of treating undiagnosed medical ailments with unprescribed drugs, being considered one of the most dynamic drug markets for pharmaceutical companies. Most experts of the domain consider that this practice will accentuate in the coming years and countries, like France,

[1] http://srvisis01.univ-mlv.fr/selfMed/

are already switching molecule sets from originally prescribed to OTC and thus opening them to the self-medication market.

Due to the increasing practice of self-medication, many government agencies are calling for the development of educational tools targeting the general public on health and medical issues. At the same time, other government agencies, usually in the same countries, are requesting to implement computerized applications based on open data. This project aims to satisfy both calls by developing a self-medication Web application fueled on Linked Open Data sets dedicated to drugs and medicine. We consider that using open data is one of the most efficient ways to collect data and knowledge on the medical domain at large scale [4]. Moreover, as explained in [5], semantic technologies will prove to be amongst the most reliable methods to search new data and knowledge into these data sets.

This project leans on the experience of implementing a similar system for the French market. Together with Pr. Jean-Paul Giroud (PhD, MD, former pharmacology WHO expert), we have contributed to the practice of a safe and efficient self-medication by releasing several books, recently [3], and web applications, one of which has been accessible for 4 years to the 6 million clients of three major insurance companies in France. Recently, an iOS application has been released and is thus accessible to everyone purchasing the Web application. Intuitively, this application aims to provide objective (i.e. not influenced by pharmaceutical companies), adapted to the general public information on drug products sold in France. In just a few interactions, one can select a symptom and obtain information on associated therapeutic classes and drugs. A main feature of this application is to rate drugs on an efficiency/tolerance ratio given the molecules contained in the drug. The cornerstone of the system is a drug database which is the result of years of medical and pharmacological research. The back office of the Web application is based on the use of a main ontology which has been designed an enriched using inductive reasoning over the drug database [2]. Moreover, this ontology serves in ensuring and enhancing the data quality of the drug database by performing some molecule oriented inferences [1].

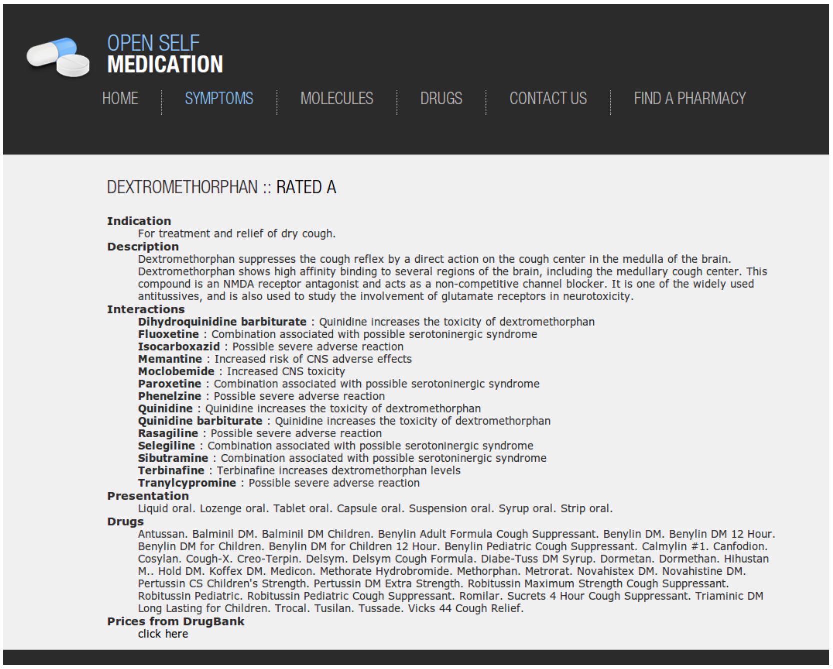
Essentially, none of the data contained in our French system are originating from the open data initiative, hence the challenge of developing one almost solely based on LOD datasets and still be of practical value to the general Web users. The main motivation of this project is to develop a self-medication application adapted to the north American market by leveraging on our experience on the French system by using open data represented in Semantic Web technologies.

2 Design choices

The design choices concerning this application were mainly motivated by the results of investigating the LOD datasets relevant to the domain of self-medication. The datasets we formally considered were *DrugBank*, *DailyMed*, *Sider*, *DBPedia*, *LinkedCT*, *Diseasome* and *FreeBase*. We finally retained only the first four since the others were not adapted in terms of items represented (e.g. *diseasome*) and medical skills required by the end-user to understand the information contained. A in-depth study of our selected datasets rapidly emphasized that the central concept of our application would not be the same as our French system. The central concept in the French system corresponds to a Drug for the following reasons: (i) a unique identifier is provided to each drug product being sold in France, (ii) information sources are available, although distributed over several providers, on these products and (iii) market evolution on these products are also available, i.e. emergence and withdrawal of products. Thus it is possible to maintain an almost exhaustive and accurate drug database for these drugs. It also enables us to rate drugs, in addition to rating molecules, which is quite useful in cases of compound drugs containing several molecules with different dosages. In the case of drugs found on LOD, we have not found an accurate and consistent list of drugs available on the market at a given time. For this reason, we currently have decided to only rate molecules and to consider non compound drugs, which generally have higher rates and better properties, i.e. less drug interactions, contraindications, side-effects.

Then comes the question of selecting the molecules fitting in self-medication? Replying to this question is not easy since we can not rely exclusively on the fact that a drug is OTC or not. For instance, in France, some drugs are OTC but we do not recommend them for self-medication due to their characteristics, e.g. contraindications. Our selection method uses our inductively developed ontology [2]. Intuitively, we have selected the most self-medication relevant clinical signs and molecules in terms of their properties, generated SPARQL queries from navigating along the subsumption hierarchy of molecule concepts and executed those queries over our RDF repository. Our first approach was based on identifying molecules with the WHO ATC classification but we rapidly found out that this identifier is sparsely used in *DrugBank* and *DBPedia*, the central dataset of our application. So we complemented our ontology with a new identifier solution based on *DrugBank* keys. This processing step was performed semi-automatically using different sources such as *DBPedia* and over classifications, e.g. *EphMRA*. At this stage, we also found out that some molecules could not be matched because they could not be found due to their non-commercialization in North America (anyway, most of them were badly rated on the French system and considered useless by many experts).

We now consider the access to those molecules from a user interface point of view. After several tentatives on the French systems, we found out that it is more convenient for the end-user to access molecules and drugs from mild clinical signs. This enabled us to design a classical self-medication scenario: the end-user selects an entry among the most frequent self-medication related symptoms, a list of adapted molecules is then displayed with short indication and its rating (from A to E). More information, e.g. contraindication, side-effects, food interactions and drugs, can be obtained from selecting a molecule. Finally, among the list of drugs, some of them, i.e. those that provide useful information and relevant in a self-medical context, can be selected to get more information, i.e. adverse events, precautions, warnings. Because we were not satisfied with the amount of drug information one can retrieve from LOD datasets, the molecule information page provides a link to *DrugBank* pages providing drug prices (Figure 1).



DEXTROMETHORPHAN :: RATED A

Indication
For treatment and relief of dry cough.

Description
Dextromethorphan suppresses the cough reflex by a direct action on the cough center in the medulla of the brain. Dextromethorphan shows high affinity binding to several regions of the brain, including the medullary cough center. This compound is an NMDA receptor antagonist and acts as a non-competitive channel blocker. It is one of the widely used antitussives, and is also used to study the involvement of glutamate receptors in neurotoxicity.

Interactions
Dihydroquinidine barbiturate : Quinidine increases the toxicity of dextromethorphan
Fluoxetine : Combination associated with possible serotoninergic syndrome
Isocarboxazid : Possible severe adverse reaction
Memantine : Increased risk of CNS adverse effects
Moclobemide : Increased CNS toxicity
Paroxetine : Combination associated with possible serotoninergic syndrome
Phenelzine : Possible severe adverse reaction
Quinidine : Quinidine increases the toxicity of dextromethorphan
Quinidine barbiturate : Quinidine increases the toxicity of dextromethorphan
Rasagiline : Possible severe adverse reaction
Selegiline : Combination associated with possible serotoninergic syndrome
Sibutramine : Combination associated with possible serotoninergic syndrome
Terbinafine : Terbinafine increases dextromethorphan levels
Tranylcypromine : Possible severe adverse reaction

Presentation
Liquid oral. Lozenge oral. Tablet oral. Capsule oral. Suspension oral. Syrup oral. Strip oral.

Drugs
Antussan. Balminil DM. Balminil DM Children. Benylin Adult Formula Cough Suppressant. Benylin DM. Benylin DM 12 Hour. Benylin DM for Children. Benylin DM for Children 12 Hour. Benylin Pediatric Cough Suppressant. Calmylin #1. Canfodion. Cosylan. Cough-X. Creo-Terpin. Delsym. Delsym Cough Formula. Diabe-Tuss DM Syrup. Dormetan. Dormethan. Hihustan M.. Hold DM. Koffex DM. Medicon. Methorate Hydrobromide. Methorphan. Metrorat. Novahistex DM. Novahistine DM. Pertussin CS Children's Strength. Pertussin DM Extra Strength. Robitussin Maximum Strength Cough Suppressant. Robitussin Pediatric. Robitussin Pediatric Cough Suppressant. Romilar. Sucrets 4 Hour Cough Suppressant. Triaminic DM Long Lasting for Children. Trocal. Tusilan. Tussade. Vicks 44 Cough Relief.

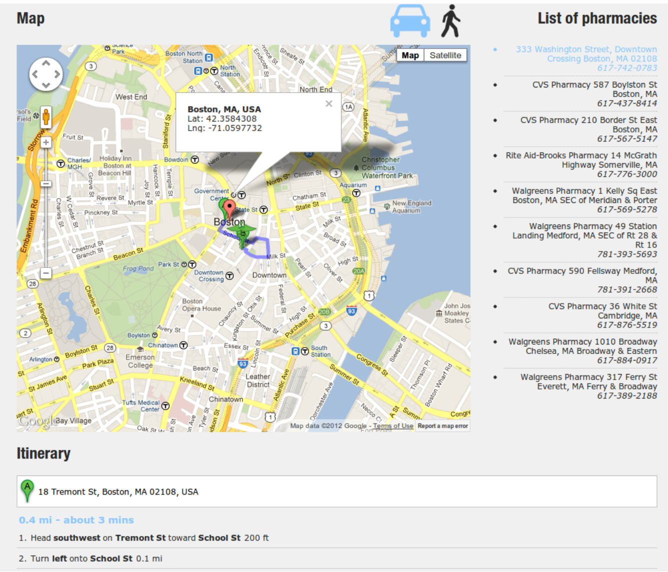
Prices from DrugBank
click here

Fig. 1. Screenshot of the Dextromethorphan molecule

In order to inform the general public on non self-medication molecules and drugs, the application also proposes to access information on all drugs and molecules obtain on the LOD datasets were using. Up to now, we do not provide ratings for these molecules since, for deontological reasons, we should not promote them in a self-medication context. Finally, to complete the self-medication service, we propose the end-user to locate the pharmacies surrounding his current or a given place. This tool suggests the ten closest pharmacies and provides driving/walking directives. For this service, we harvested the web for pharmacy coordinates in France and the US.

3 Implementation details

The Web application uses HTML5 and CSS3 for the user interface and programs are written using PHP, Javascript and its JQuery library. The geolocation page (Figure 2) uses the Google Map API version 3 and the navigator HTML5 extension. Concerning the LOD datasets, we are not using SPARQL endpoints due their relative unreliability[2]. Instead, we loaded some dumps and put them on our own OpenRDF Sesame triple store. We are using the phpSesame PHP Client library which enables to access OpenRDF's Sesame Framework via HTTP requests. All non RDF data, i.e. list of mild clinical signs, molecules and their ratings as well as pharmacy coordinates, are stored in a Mysql database instance. A stored procedure enables to compute the geodesic distance between a given latitude/longitude pair and coordinates of the stored pharamacy coordinates. Retrieving data from these datastores is performed using REST services which are using SPARQL and SQL queries. We are aiming to propose an API for these services in order to enable interested developers to access them directly.



Map

Boston, MA, USA
Lat: 42.3584308
Lng: -71.0597732

List of pharmacies

- 333 Washington Street, Downtown Crossing Boston, MA 02108 *617-742-0783*
- CVS Pharmacy 587 Boylston St Boston, MA *617-437-8414*
- CVS Pharmacy 210 Border St East Boston, MA *617-567-5147*
- Rite Aid-Brooks Pharmacy 14 McGrath Highway Somerville, MA *617-776-3000*
- Walgreens Pharmacy 1 Kelly Sq East Boston, MA SEC of Meridian & Porter *617-569-5278*
- Walgreens Pharmacy 49 Station Landing Medford, MA SEC of Rt 28 & Rt 16 *781-393-5693*
- CVS Pharmacy 590 Fellsway Medford, MA *781-391-2668*
- CVS Pharmacy 36 White St Cambridge, MA *617-876-5519*
- Walgreens Pharmacy 1010 Broadway Chelsea, MA Broadway & Eastern *617-884-0917*
- Walgreens Pharmacy 317 Ferry St Everett, MA Ferry & Broadway *617-389-2188*

Itinerary

A 18 Tremont St, Boston, MA 02108, USA

0.4 mi - about 3 mins

1. Head **southwest** on **Tremont St** toward **School St** 200 ft
2. Turn **left** onto **School St** 0.1 mi

Fig. 2. Screenshot of the geolocation page

4 Lessons learned

The development of this application emphasizes the heterogeneity and data quality issues of the data contained in some LOD datasets. In terms of heterogeneity, some molecules are lacking some important properties, e.g. *ATC* or *DrugBank* molecule identifiers, and are not properly linked to other datasets. Although some drug products are well documented, others are totally absent. At the same time, some important properties are missing for some popular drugs.

It was not a surprise to discover that the content on diseases, molecules and drugs on datasets such as LOD are mainly targeting the community of health

[2] http://labs.mondeca.com/sparqlEndpointsStatus/index.html

care professionals. Studying them, we found out that some information are still comprehensible for the general public.

Finally, to the best of our knowledge, there does not exist a free, open source repository of addresses of health care actors and care facilities. Such datasets are highly desirable for developers aiming to implement applications providing directives to hospitals, clinics, pharmacies, etc.

5 Conclusion and Future works

We consider that this application already provides useful information to the general public on mild clinical signs and self-medication molecules. Nevertheless, we were disappointed with the correctness and completeness of drug information. In order to deal with these issues, we have decided to show only those drugs that satisfy a certain quality, i.e. given in terms of correspondences with our own ontology. An unavoidable task for such a system is to update the datasets in our repository, i.e. we aim to periodically load dumps of the used datasets We also hope to get some end-user feedbacks to improve the system and develop new features.

In terms of future work, we consider several directions. A first one is concerned with implementing a similar tool for health care professionals since the content of the LOD datasets are more oriented to their skills. We already know that many general practitioners in France are using either the books or applications we have produced on self-medication. Pursuing in the general public direction, we are aiming to develop a multilingual version and to adapt the terms used in some properties of these datasets such that they are not idiosyncratic to medicine.

6 Acknowledgment

I would like to thank Pr. Jean-Paul Giroud for his medical expertise and support in implementing this application. I would also like to thank the following Université of Paris Est Marne la Vallée Master students for their contribution in implementing a preliminary version: Yann Bougeoisat, Yan De Backer, Jacques Nicolas De Lamballerie, Rafarandramasy Razafindramanana.

References

1. O. Curé, *Improving the data quality of drug databases using conditional dependencies and ontologies*, ACM Journal of Data and Information Quality, (Accepted for publication in 2012).
2. O. Curé and J.-P. Giroud, *Ontology-based data quality enhancement for drug databases*, in Proc. of Int Workshop on Health Care and Life Sciences Data Integration for the Semantic Web, WWW Conference.
3. J.-P. Giroud, *Les médicaments sans ordonnance, les bons et les mauvais!*, Editions de la Martinière, France, 2011.

4. C. Jonquet, P. LePendu, S. M. Falconer, A. Coulet, N. F. Noy, M. A. Musen, and N. H. Shah, *Ncbo resource index: Ontology-based search and mining of biomedical resources*, J. Web Sem., 9 (2011), pp. 316–324.
5. P. LePendu, M. A. Musen, and N. H. Shah, *The age of data-driven medicine: Mining the electronic health record*, in ICBO, 2011.
6. OECD, *Health at a Glance 2011: OECD Indicators*, OECD Publishing, 2011.

Appendix: Compliance with Semantic Web Challenge Requirements

Compliance with minimal requirement

✓ *The application has to be an end-user application*
By definition, a self-medication application targets the general public. We have designed a user-friendly interface that enables to obtain drug information within 4 steps.
✓ *Information sources used should be under diverse ownership or control*
The four information sources used in this project are under different controls: LOD datasets, data extracted from the Web on drug store locations, our self-medical ontology and molecule ratings which are our own.
✓ *Information sources should be heterogeneous*
Several data formats have been used on the project: RDF data from different LOD datasets, CSV and XML concerning drug store geolocation, OWL for our ontology and relational data for molecule ratings and symptoms.
✓ *Information sources used should contain substantial quantities of real world data*
We have collected around 20.000 drug store locations (including address, latitude and longitude), our ontology contains more than 5.000 molecules and the amount of triples in our datasets are close to 2 million triples. All are real world data.
✓ *Meaning must be represented using Semantic Web technologies*
An OWL molecule ontology supports the selection of molecules considered in the application. All molecules and drugs information are represented in RDF. An OWL reasoner and SPARQL queries are used through the process of molecule selection and data retrieval.
✓ *Data must be manipulated/processed in interesting ways*
Section 3 presents some details on retrieving the most informative and understandable for the general public data from our selected LOD datasets.
✓ *Semantic information processing has to play a central role in achieving things that alternative technologies cannot do as well, or at all*
Reasoning over the semantics attached to a graph of molecules would not be possible with other technologies. The complete stack of technologies: ontology languages, triple stores, query language and reasoners are absent or not as mature in alternative technologies, e.g. relational or graph databases.

Compliance with additional desirable features

✓ *The application provides an attractive and functional Web interface*
With a 4 steps approach, the end-user selects a symptom, obtains a molecule set, gets information on a molecule its drugs. Moreover, he can localize a nearby drug store and gets driving or walking directives. The application is implemented using HTML5/CSS3 application and hence enables usage on a Web connected terminal (tablets, smartphones).
✓ *The application should be scalable*
The application uses all data that is currently relevant to the self-medication domain on the Semantic Web. The architecture of the application is able to process larger data, e.g. different sources for adverse events, but the domain surrounding self-medication is finite and of a limited size.
✓ *Rigorous evaluations have taken place that demonstrate the benefits of semantic technologies, or validate the results obtained.*
The quality of the information displayed to the end-user is checked using our molecule ontology and a set of mappings to standard terminologies.
✓ *Novelty, in applying semantic technology to a domain or task that have not been considered before*
To the best of our knowledge, we do not know of any self-medication application using semantic technologies and open datasets.
✓ *Functionality is different from or goes beyond pure information retrieval not been considered before*
By using the semantics that is encoded in our molecule ontology, our application goes beyond traditional information retrieval.
✓ *The application has clear commercial potential*
The french, non open data self-medication Web application that inspired this project has been commercialized 4 years ago. Section 1 clearly motivates the design of such application.
✓ *Contextual information is used for ratings or rankings*
We are ranking source of side-effects, i.e. we prefer Sider and DailyMed information when available to some information found on DrugBank or DBpedia.
✓ *Multimedia documents are used in some way*
Geolocation of a nearby drug store is the most advanced form of multimedia used in this application. It provides HTML5 geolocation and driving/walking directives to get to one of 10 closest drugstores.
There is a use of dynamic data (e.g. workflows), perhaps in combination with static information
There is no use of dynamic data such as workflow or data streams. Up to now the domain does not apply to such use.
✓ *The results should be as accurate as possible*
Drugs with too many missing values are rejected and not presented to the end-user.
There is support for multiple languages and accessibility on a range of devices
The use of HTML5 supports the use of the Web application on all devices connected to the Web. The support of multiple languages has not be addressed although the used datasets permits it. It would require to that medical experts in each language, e.g. chinese, japanese.