Semi-automatic loading of a microbial risk in food database thanks to an ontology in the context of Linked Data

Buche P.[1], Dervaux S.[2], Dibie-Barthélemy J.[2,3], Ibanescu L.[2,3], Hecht T. [2], Soler L.[2]

[1] *INRA - UMR IATE, 2, place Pierre Viala, F-34060 Montpellier Cedex 2, and LIRMM, Montpellier, France*

[2] *INRA - Unité Mét@risk, 16 rue Claude Bernard, F-75231 Paris Cedex 5, France*

[3]*AgroParisTech - 16 rue Claude Bernard, F-75231 Paris Cedex 5, France*

OBJECTIVE(S)

A preliminary step in microbial risk assessment in food is to gather and to capitalize experimental data in order to feed simulation models. We have designed, in the predictive modeling platform Sym'Previus (www.symprevius.org), a relational database, MicroRisk-RDB [1], to store experimental data in microbial risk. Data capitalization task encounters a challenging lock. Indeed, original data are spread out in heterogeneous data sources (scientific papers, technical reports / sheets, PhD thesis …). Moreover, they are expressed in heterogeneous formats (mainly tables, text and graphics) and vocabularies. Manual entering of data in a database (e.g. MicroRisk-RDB) is therefore a time-consuming task. We present in this paper methods and tools to facilitate data capitalization and reusability in the framework of the Linked Data initiative[1], which promotes the publication and connection of structured data on the Web.

METHODS

We are currently designing @Web-Tool[2], a semi-automatic tool guided by an ontology, to help domain experts to populate a database (e.g. MicroRisk-RDB) with data tables found in scientific publications, especially on the Web. The notion of 'ontology' is one of the key-stone of the Linked Data initiative and more generally of the Semantic Web. It is a vocabulary used to organize the knowledge of a domain of interest, so that data can be exported, translated, queried, and unified across independently developed systems and services[3]. In @Web-Tool, the user first downloads an HTML scientific document, and then data tables are semi-automatically identified and extracted. A graphical user-friendly interface helps the user to annotate data tables thanks to the same ontology used to index the database to be populated. The user may manually suggest new candidate terms to be added in the ontology which are managed by the administrator. @Web-Tool has been designed in the framework of the Linked Data initiative using the W3C standards stack [2].

RESULTS

@Web-Tool is currently tested by different experts in the framework of the French ANR project MAP'OPT (Equilibrium Gas Composition in Modified Atmosphere Packaging and Food Quality) in order to collect data on food packaging permeabilities used to develop new growth simulation model taking into account modified atmosphere conditions. Our tests rely on the MAP'OPT-Onto ontology [3], which has been built as an extension of the ontology used to index MicroRisk-RDB. Current version

---

[1] http://linkeddata.org/

[2] http://www.paris.inra.fr/metarisk/research_unit/knowledge_engineering/softwares/web__1

[3] http://tomgruber.org/writing/ontology-definition-2007.htm

of MAP'OPT-Onto is composed of 1103 concepts representing domain objects (food products, microorganisms, and packagings), 23 quantity concepts and 83 associated units of measure. Finally 20 relation concepts represent the interdependence between domain objects and quantities (e.g. Temperature_Cardinal_Value relation concept express the relation between a given microorganism and its temperature conditions of growth).

CONCLUSIONS AND IMPACT OF THE STUDY

@Web-Tool is designed as a microbial risk in food data semantic Web annotation tool allowing a better reusability of experimental data in the framework of the Linked Data initiative. Our aim is eventually to publish MAP'OPT-Onto on the Web in order to set up a standardized vocabulary in microbial risk in food. In the very next future, we will link our MAP'OPT-Onto with international standardized ontologies, such as AGROVOC[4], the FAO thesaurus for agriculture, OM [4] and QUDT [5], in order to promote international data sources interconnection. Linking our MAP'OPT-Onto with other ontologies can provide us a richer definition of concepts we manipulate, more experimental data, access to conversion tools of OM and, so, a refined way of annotating and reusing the knowledge gathered. New bridges will therefore be established between data that will benefit to the simulation models. Finally, future evolutions of @Web-Tool will enhance data manual annotation efficiency thanks to semi-automatic annotation algorithms of data tables [6, 7].

REFERENCES

[1] *Leporq B., Membré J.-M., Dervin C., Buche P. and Guyonnet J.P. (2005). The ``Sym'Previus'' software, a tool to support decisions to the foodstuff safety . International Journal of Food Microbiology, 100(1-3) 231-237.*

[2] *Berners-Lee T., Hendler J., and Lassila O. (2001) « The semantic web », Sci. Am., vol. 284, n° 5, p. 28–3. Also http://www.w3.org/2001/sw/.*

[3] *Touhami R., Buche P., Dibie-Barthélemy J., and Ibanescu L. (2011) « An ontological and terminological resource for n-ary relation annotation in web data tables », in On the Move to Meaningful Internet Systems: OTM 2011, Springer, p. 662–679.*

[4] *Rijgersberg H., Wigham M., and Top J.L. (2011) « How semantics can improve engineering processes: A case of units of measure and quantities », Adv. Eng. Informatics, vol. 25, n° 2, p. 276–28.*

[5] *R. Hodgson et P. J. Keller, QUDT-Quantities, Units, Dimensions and Data Types in OWL and XML. 2011.On line document available at http://www.qudt.org/*

[6] *G. Hignette, P. Buche, O. Couvert, J. Dibie-Barthélemy, D. Doussot, O. Haemmerlé, E. Mettler, L. Soler (2008). Semantic annotation of Web data applied to risk in food. International Journal of Food Microbiology, vol. 128(1), Elsevier, pp. 174-180.*

[7] *Buche P., Dibie-Barthélemy J., Ibanescu L. and Soler L. (2013) Fuzzy Web Data Tables Integration Guided by an Ontological and Terminological Resource. IEEE Trans. Knowl. Data Eng. 25(4): 805-819.*

[4] http://aims.fao.org/standards/agrovoc