



**HAL**  
open science

## Flexible Querying of Web data to Simulate Bacterial Growth in Food

Patrice Buche, Olivier Couvert, Juliette Dibie-Barthelemy, Gaëlle Hignette,  
Eric Mettler, Lydie Soler

► **To cite this version:**

Patrice Buche, Olivier Couvert, Juliette Dibie-Barthelemy, Gaëlle Hignette, Eric Mettler, et al.. Flexible Querying of Web data to Simulate Bacterial Growth in Food. Food Microbiology, 2011, 28 (4), pp.685-693. 10.1016/j.fm.2010.07.002 . lirmm-00538961

**HAL Id: lirmm-00538961**

**<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00538961>**

Submitted on 29 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18

# Flexible querying of Web data to simulate bacterial growth in food

Patrice Buche<sup>\*1,7,8</sup>, Olivier Couvert<sup>2,3,7</sup>, Juliette Dibie-Barthélemy<sup>4,5</sup>, Gaëlle Hignette<sup>4,5</sup>,  
Eric Mettler<sup>6,7</sup>, Lydie Soler<sup>4</sup>

<sup>1</sup>INRA-UMR IATE, 2 place Viala, 34060 Montpellier Cedex 2, France  
E-mail: buche@supagro.inra.fr

<sup>2</sup>ADRIA Développement, Creac'h Gwen, 29196 Quimper Cedex, France

<sup>3</sup>Université Européenne de Bretagne, Université de Brest, LUBEM EA 3882 - UMT  
Physiopt 08.3, 6 rue de l'Université, F-29334 Quimper Cedex, France

<sup>4</sup>INRA Mét@risk 16, rue Claude Bernard 75231 Paris Cédex 5 France

<sup>5</sup>AgroParisTech, UFR Informatique, 16, rue Claude Bernard, 75 231 Paris Cedex 05, France

<sup>6</sup>Soredab (Groupe SOPARIND BONGRAIN), La Tremblaye,  
78125 La Boissière-Ecole, France

<sup>7</sup>Groupement d'Intérêt Scientifique Sym'Previus, 147 rue de l'Université,  
F-75007 Paris, France

<sup>8</sup>LIRMM, CNRS-UM2, F-34392 Montpellier, France

---

<sup>1</sup> corresponding author

Version postprint

19 **Abstract**

20 A preliminary step in microbial risk assessment in foods is the gathering of experimental data.  
21 In the framework of the [Sym'Previus project](#), we have designed a complete data integration  
22 system opened on the Web which allows a local database to be complemented by data  
23 extracted from the Web and annotated using a domain ontology. We focus on the Web data  
24 tables as they contain, in general, a synthesis of data published in the documents. We propose  
25 in this paper a flexible querying system using the domain ontology to scan simultaneously  
26 local and Web data, this in order to feed the predictive modeling tools available on the  
27 Sym'Previus platform. Special attention is paid on the way fuzzy annotations associated with  
28 Web data are taken into account in the querying process, which is an important and original  
29 contribution of the proposed system.

30 **Keywords :** Web data, flexible querying, ontology, predictive microbiology

31

32

## 33 Introduction

34 A preliminary step in microbial risk assessment in foods is the gathering of experimental data  
35 (Tamplin *et al.* 2003, Baranyi and Tamplin 2004, McMeekin *et al.* 2006).

36 In the framework of the Sym'Previus project (Couvert *et al.* 2007 and  
37 <http://www.symprevius.org>), we have designed a complete data integration system opened on  
38 the Web which allows a local database (Buche *et al.* 2005) to be complemented by data  
39 extracted from the Web (Hignette *et al.* 2008). The local data were classified by means of a  
40 predefined vocabulary organized in taxonomy, called ontology. This ontology is used to  
41 extract pertinent data from the Web. We focus on the Web data tables as they contain, in  
42 general, a synthesis of data published in the documents. Our aim is to integrate the data tables  
43 found on the Web with the local data by means of a flexible querying system which allows the  
44 end-user to retrieve the nearest local and Web data corresponding to his/her selection criteria.  
45 With our solution, the end-user may simultaneously and uniformly query local and Web data  
46 in order to feed the predictive modeling tools available on the Sym'Previus platform.

47 These developments have been introduced in the predictive modeling program Sym'Previus  
48 ([www.symprevius.org](http://www.symprevius.org)). Actually, to take into account the food matrix effect, predictive  
49 models need raw data obtained from food product. Considering the large diversity of foods, a  
50 local database seems to be too limited (i) to gather information for all food products, and (ii)  
51 to have enough and adequate data to take into account the food variability. The simultaneous  
52 querying in local and web data increases the accuracy and the pertinence of the simulation  
53 results.

54 We first remind the semi-automatic annotation method (implemented in the @WEB tool, see  
55 [@Web demo](#)) which allows data to be retrieved from data tables found in scientific  
56 documents on the Web and to be annotated thanks to the ontology. As the local data and the

57 Web data tables were all together indexed by the ontology, it is therefore possible to use the  
58 terminology defined in the ontology in order to query simultaneously those two sources of  
59 information. Second, we present the original contribution of the paper, which consists in the  
60 design of the flexible querying system, called MIEL++. This system allows the end-user to  
61 query simultaneously and in a transparent way the local data and the semantic annotated Web  
62 data, thanks to the ontology. It is flexible because (i) it allows the end-user to express  
63 preferences in his/her selection criteria and (ii) it takes into account, in the answers content,  
64 the different kinds of fuzziness of the semantic annotated Web data. This second point is  
65 essential to deal with the uncertainty of the Web data and with the imperfection of their  
66 annotations. Third and finally, experimental results are presented and discussed.

67

## 68 **Materials and methods**

69 Our annotation method which allows the Web data tables to be indexed thanks to the  
70 vocabulary defined in the ontology has already been presented in details by Hignette *et al.*  
71 (2008). It is briefly recalled in the first paragraph of this section. The content of the Web data  
72 tables must be indexed according to the ontology in order to be queried. This indexation  
73 associates a set of annotation graphs with each row of a Web data table. This method is  
74 presented in the second paragraph. Then, we present in the third paragraph, the automatic  
75 querying method which uses the index associated with the Web data tables in order to perform  
76 the MIEL++ query. Finally, we present in the fourth paragraph, the way the experimental data,  
77 extracted thanks to the MIEL++ querying system, are used to estimate the parameters of the  
78 simulation model.

79

### 80 **Automatic annotation method of a Web data table**

81 Web data tables are semi-automatically annotated by means of a predefined vocabulary, called  
82 ontology (see definition in Table 1). This ontology is composed of data types meaningful in  
83 the domain of risk in food and semantic relations linking those data types. The structure of the  
84 ontology is presented in Figure 1. Data types are described in two different ways depending on  
85 whether their associated values are symbolic (*Food product*, *Microorganism* ...) or numeric  
86 (*Temperature*, *pH* ...). Symbolic types are described by taxonomies of possible values (for  
87 example, a taxonomy of microorganisms). The taxonomy of possible values associated with a  
88 symbolic type defines its domain of values. Numeric types are described by their possible set  
89 of units (for example, °C or °F for *Temperature*, but no unit for *pH* or *a<sub>w</sub>*), and their possible  
90 numeric range (for example, [0, 14] for *pH*). The numeric range associated with a numeric  
91 type defines its domain of values. Semantic relations (see definition in Table 1) are defined by  
92 their signature which is composed of a result data type and a set of access data types. For  
93 example, the relation *GrowthParameterAw*, representing the growth limits of a  
94 microorganism for any food product, has for access type the symbolic type *Microorganism*  
95 and for result type the numeric type *a<sub>w</sub>*. Our annotation method first annotates the symbolic  
96 columns and the numeric columns and then uses these annotations to recognise the semantic  
97 relations present in the Web data tables (see Hignette *et al.* 2008 for more details).

98 Example: We consider a table having for legend “Reported prevalence of Campylobacter” and  
99 which is composed of two columns having respectively for title: “Product” and “Positive for  
100 Campylobacter (%)”. The first row of this table is composed of the term “Chicken products”  
101 in the cell corresponding to the “Product” column and 0.07 in the cell corresponding to the  
102 “Positive for Campylobacter (%)” column. When annotating [this table](#), the method finds that  
103 the first column is symbolic and the second one is numeric. Concerning the first column, the  
104 method annotates it by the symbolic type *Food product*. The second column is annotated by  
105 the numeric type *Samples Positive*. Finally the whole table is automatically annotated by the

106 *Prevalence* semantic relation.

107 In the following, we explain how the semantic relations used to annotate a Web data table are  
108 instanciated for each row of the Web data table in order to index it, this indexation being a  
109 preliminary step to the flexible querying process.

110

### 111 **Instanciation of a semantic relation in a Web data table into a RDF graph**

112 Once a Web data table has been annotated by one or several semantic relations, it is indexed  
113 by instances of these relations which are associated with each row of the Web data table. The  
114 instanciation (see definition in Table 1) of a semantic relation in a Web data table is  
115 represented, for each row of the table, as a Resource Description Framework (RDF) graph.  
116 RDF is the language recommended by the W3C (World Wide Web consortium) to represent  
117 semantic annotations associated with Web resources. An instance of a semantic relation  
118 associated with a row of a Web data table is composed of the instances of the result data type  
119 and the access data types of its signature which are associated with the data present in the cells  
120 of the row. The generated instanciations are fuzzy: they allow one to take into account the  
121 imprecision of the initial data in the table (for example an interval for a numeric type), the  
122 similarity comparison between the vocabulary used in the table with the vocabulary of the  
123 ontology, and the uncertainty of the annotation of the table by semantic relations. We first  
124 present briefly the theory of fuzzy sets that we use in our instanciation method, then we  
125 present how we instanciate numeric types, symbolic types and relations.

126

#### 127 *Fuzzy sets:*

128 We use the definition of fuzzy sets given by Zadeh, 1965 and Zadeh, 1978. The notion of  
129 fuzzy set is an extension of classical subsets. In the classical case, elements of a reference set  
130 X which have some properties belong to a subset A, and elements which do not have these

131 properties belong to the complementary subset of  $A$  in  $X$ . In a fuzzy set, elements can belong  
 132 partially to the fuzzy set, with a membership degree included between 0 (element which is not  
 133 part of the fuzzy set) and 1 (element which is completely part of the fuzzy set). The  
 134 membership degree of an element  $x$  of the reference set  $X$  for the fuzzy set  $A$  is denoted  $\mu_A(x)$ .  
 135 When  $X$  is defined on a continuous domain, we talk about a continuous fuzzy set; when  $X$  is  
 136 defined on a discrete domain, we talk about a discrete fuzzy set. The support of a fuzzy set  $A$   
 137 defined on a reference set  $X$  is the set (in the classical definition) of elements  $x$  of  $X$  such that  
 138  $\mu_A(x) > 0$ . The kernel of a fuzzy set  $A$  defined on a reference set  $X$  is the set (in the classical  
 139 definition) of elements  $x$  of  $X$  such that  $\mu_A(x) = 1$ .

140 A trapezoid fuzzy set TFS is a special continuous fuzzy set which is described only by its  
 141 support  $\text{sup} = [\min_{\text{sup}}, \max_{\text{sup}}]$  and its kernel  $\text{ker} = [\min_{\text{ker}}, \max_{\text{ker}}]$ . The membership degree of  
 142 a numeric value  $x$  in the reference set is then defined as follows:

143 – if  $x \leq \min_{\text{sup}}$  or  $x \geq \max_{\text{sup}}$  then  $\mu_{\text{TFS}}(x) = 0$ ;

144 – if  $\min_{\text{ker}} \leq x \leq \max_{\text{ker}}$  then  $\mu_{\text{TFS}}(x) = 1$ ;

145 – if  $\min_{\text{sup}} \leq x \leq \min_{\text{ker}}$  then  $\mu_{\text{TFS}}(x) = \frac{x - \min_{\text{sup}}}{\min_{\text{ker}} - \min_{\text{sup}}}$  ;

146 – if  $\max_{\text{ker}} \leq x \leq \max_{\text{sup}}$  then  $\mu_{\text{TFS}}(x) = \frac{x - \max_{\text{ker}}}{\max_{\text{sup}} - \max_{\text{ker}}}$  ;

147 There are several semantics for fuzzy sets, defined in Dubois and Prade, 1997:

- 148 • Preferences: the elements with the higher membership degrees are the preferred  
 149 elements. This is used in the MIEL++ querying system by the user to define query  
 150 preferences;
- 151 • Uncertainty or imprecision: there exists a “true” value, but we do not know it. The  
 152 higher is the membership degree of a value  $x$ , the more possible is  $x$  to be the “true”  
 153 value. This is used in our instantiation method to represent the imprecision of the  
 154 original data in the tables (in the instantiation of numeric types);



- 155       • Similarity: a new value is represented by its similarity with known values. The higher  
156       is the membership degree of a known value  $x$ , the more it is similar to the new value.  
157       This is used in our instantiation method to represent the similarities between a term  
158       from the web and terms from the ontology (in the instantiation of symbolic types).

159

### 160 *Instantiation of numeric types*

161 Let us consider that a Web data table has been annotated by a semantic relation. There are  
162 three possibilities in order to instantiate (see definition in Table 1) a numeric type  $t$  of the  
163 signature of the relation in a given row of the Web data table:

164 1. There is one column in the table (thus one cell in the row to annotate) that was annotated  
165 by the numeric type  $t$ . In this case, the values in the cell are used to instantiate the type: it can  
166 be an isolated value, an enumeration of isolated values, an interval or a mean with a standard  
167 error. Intervals and mean with standard error are recognised using specific patterns; if those  
168 patterns are not recognised, then all numeric values in the cell are considered as isolated.

169 2. There are several columns in the table that were annotated by the numeric type  $t$ . In this  
170 case, we have to find the relations between the columns: it is done by looking for keywords in  
171 the columns titles. A column can represent a minimum value, a maximum value or an  
172 optimum value (included between the minimum and maximum); it can also represent a mean  
173 value or a standard error.

174 3. There is no column in the table that was annotated by the numeric type  $t$ . If the numeric  
175 type  $t$  has a defined unit, we search for occurrences of a numeric value followed by this unit, in  
176 the table title or in the columns titles: those occurrences are then considered as isolated values.

177 An instance of a numeric type is represented by a continuous fuzzy set of which the reference  
178 set is the numeric range of the numeric type defined in the ontology. This fuzzy set is built  
179 from trapezoid fuzzy sets, each being created as follows:

- 180 • when recognising an isolated value  $x$  in the table, we construct a trapezoid fuzzy set  
181 with  $\text{sup} = \text{ker} = [x, x]$ ;
- 182 • when recognising an interval  $[a, b]$  in the table, either in one cell or when  $a$  is the value  
183 in a cell recognised as minimum and  $b$  is the value in a cell recognised as maximum  
184 with no cell recognised as optimum, we construct a trapezoid fuzzy set with  $\text{sup} = \text{ker}$   
185  $= [a, b]$ ;
- 186 • when recognising a cell as minimum, its minimum numeric value being  $\text{min}$ , a cell as  
187 maximum, its maximum numeric value being  $\text{max}$  and a cell as optimum, its values  
188 being included in the minimum interval  $[a, b]$ , we construct a trapezoid fuzzy set with  
189  $\text{sup} = [\text{min}, \text{max}]$  and  $\text{ker} = [a, b]$ ;
- 190 • when recognising a mean  $m$  and a standard error  $e$ , we construct a trapezoid fuzzy set  
191 with  $\text{sup} = [m - e, m + e]$  and  $\text{ker} = [m, m]$ .

192 Once all trapezoid fuzzy sets have been created, the instantiation of the numeric type is the  
193 union of all those sets (for example, there can be a union of several isolated values).

194

195 Example: Table 2 is annotated with the semantic relation of the ontology

196 *GrowthParameterAw*, with the access type *Microorganism* and the result type  $a_w$ . This result  
197 type is instantiated, for the first row of the data table, as a unique trapezoid fuzzy set with  $\text{sup}$   
198  $= [0.943, 0.97]$  and  $\text{ker} = [0.95, 0.96]$ . This fuzzy set is represented in Figure 2.

199

### 200 *Instantiation of symbolic types*

201 Let us consider that a Web data table has been annotated by a semantic relation. In order to  
202 instantiate (see definition in Table 1) a symbolic type  $t$  of the signature of the relation in a  
203 given row of the Web data table, we construct a discrete fuzzy set. The reference set of this  
204 fuzzy set is the taxonomy of possible values of the symbolic type  $t$ . The membership degree of

205 a term  $x$  of the ontology in the fuzzy set is the term similarity between  $x$  and the term in the  
206 cell that was annotated by the symbolic type  $t$ . In the corpus of tables we used for  
207 experimentations, it did not happen that several columns in a table were annotated by the  
208 same symbolic type  $t$ , however, would that happen, we would construct a union of fuzzy sets  
209 (one fuzzy set for each column).

210 The term similarity between a term  $x$  of the ontology and the term in the cell has already been  
211 presented in Hignette *et al.* (2008) and is briefly recalled in the following. Both terms are  
212 transformed into weighted vectors: the coordinate system of the vectors is the set of all  
213 possible words (i.e. all words in the ontology plus the words of the terms to compare with the  
214 ontology), the coordinate values associated with a given vector represent the weight of those  
215 words in the term (1 if the word is present in the term, 0 otherwise). The similarity between  
216 both terms is computed as the cosine similarity measure between the two weighted vectors,  
217 which is one of the most popular similarity measures described by Lin, 1998.

218 Example: Table 2 is annotated with the semantic relation of the ontology  
219 *GrowthParameterAw*, with the access type *Microorganism* and the result type *aw*. The access  
220 type is instantiated, for the first row of the data table, as a unique fuzzy set, defined as  
221 follows:  $\{0.5/Clostridium\ perfringens, 0.5/Clostridium\ botulinum\}$  which means that the  
222 term, *Clostridium*, in the cell is similar to the terms *Clostridium perfringens* and *Clostridium*  
223 *botulinum* of the ontology with a similarity measure of 0.5. If other *Clostridium* are defined in  
224 the ontology, they will also appear in the fuzzy set.

225

### 226 *Instantiation of semantic relations*

227 Once all the types of the signature of a semantic relation have been instantiated for a given  
228 row of a Web data table, we can instantiate the semantic relation for this row: we create an  
229 instance of the relation which is composed of the instances of the numeric and symbolic types

230 of its signature which were created for the row.

231 Example: Let us consider that Table 2 has been annotated by the semantic relation

232 *GrowthParameterAw*. Figure 3 shows the RDF graph which represents the instantiation of

233 this relation in the first data row of Table 2. In Figure 3, the RDF graph expresses that the row

234 (having the identifier *uriRow1* in the RDF graph) is annotated by a discrete fuzzy set, called

235 *DFSRI*. This fuzzy set has a semantic of similarity and indicates the list of the closest

236 semantic relations of the ontology used to annotate the first row. Only the semantic relation

237 *GrowthParameterAw* belongs to this fuzzy set with the pertinence score of 1.0, which

238 expresses the degree of certainty associated with the semantic relation recognition by our

239 annotation method. The access type of the relation, which is an instance of the symbolic type

240 *Microorganism*, is instantiated by a discrete fuzzy set, called *DFSI*. This fuzzy set has a

241 semantic of similarity and indicates the list of the closest terms of the ontology compared with

242 the term *Clostridium*. Two terms (*Clostridium Perfringens* and *Clostridium Botulinum*)

243 belong to this fuzzy set with a membership degree of 0.5. The result type of the relation,

244 which is an instance of the numeric type *aw*, is instantiated by a continuous fuzzy set, called

245 *CFSI*. This fuzzy set has a semantic of imprecision and indicates the possible growth limits

246 ([0.943, 0.97]) and the possible optimal growth limits ([0.95, 0.96]).

247 In the following, we call annotations of a Web data table the instantiations of the semantic

248 relations which have been recognised in the table.

249

### 250 **Simultaneous flexible querying of the RDF graph database and the local database**

251 The MIEL++ querying system relies on the domain ontology used to index the local data and

252 the Web data tables. MIEL++ allows the end-user to retrieve the nearest local and Web data

253 corresponding to his/her selection criteria expressed as fuzzy sets and representing his/her

254 preferences. The ontology -more precisely the taxonomies of values associated with symbolic

255 types- is used in order to assess which data can be considered as “near” to the user’s selection  
256 criteria. A query is asked to the MIEL++ system through a single graphical user interface,  
257 which relies on the domain ontology. The query is translated into the query language of each  
258 data source: an SQL query in the relational local database (see Buche *et al.*. 2005 for more  
259 details) and a SPARQL query in the RDF graph base. SPARQL is the querying language  
260 recommended by the W3C to query annotations expressed in RDF graphs. The global answer  
261 to the query is then the union of the local answers in each data source, which are ordered  
262 according to their relevance to the query selection criteria. In this paper, we focus on three  
263 original aspects of the SPARQL querying: (1) the use of the taxonomies of values associated  
264 with the symbolic types to enlarge the querying, (2) the way comparisons between the user’s  
265 selection criteria and fuzzy annotations of Web data tables are done, (3) the total order on the  
266 answers defined to retrieve the most pertinent data to the user.

267 Example: Let us consider a MIEL++ query  $Q$  expressed in the relation *GrowthParameterAw*  
268 and having for selection criteria ( $aw=awPreference$ ) and  
269 (*Microorganism=MicroPreferences*). The continuous fuzzy set *awPreferences*, which is equal  
270 to [0.9, 0.94, 0.97, 0.99], means that the end-user is first interested in  $a_w$  values in the interval  
271 [0.94, 0.97], but he/she accepts to enlarge the querying till the interval [0.9, 0.99]. The  
272 discrete fuzzy set *MicroPreferences*, which is equal to {1.0/*Gram+*, 0.5/*Gram-*}, means that  
273 the end-user is interested in microorganisms which are first *Gram+* and then *Gram-*. This  
274 fuzzy set defines implicitly user’s preferences for microorganisms which are kinds of *Gram+*  
275 and *Gram-*. Besides, the taxonomy of values associated with the symbolic type  
276 *Microorganism* contains the terms *Clostridium Botulinum*, *Clostridium Perfringens* and  
277 *Staphylococcus Spp.* which are kind of *Gram+* and *Salmonella spp.* which is a kind of *Gram-*.  
278 In order to take those implicit preferences into account in the querying, we propose to perform  
279 a closure of the fuzzy set *MicroPreferences* (see Thomopoulos *et al.* 2003 and Thomopoulos

280 *et al.* 2006 for more details). Intuitively, the closure propagates degrees of preferences to more  
281 specific values of the taxonomy. By example, the closure of the fuzzy set *MicroPreferences*  
282 is: {1.0/*Gram+*, 0.5/*Gram-*, 1.0/*Clostridium Botulinum*, 1.0/*Clostridium Perfringens*, 1.0/  
283 *Staphylococcus Spp.*, 0.5/*Salmonella*}.

284 In order to build the answer to a query, selection criteria representing user's preferences  
285 expressed as fuzzy sets must be compared with fuzzy annotations, which are associated with  
286 the Web data tables to query. But, as we saw previously, the fuzzy sets used in the annotations  
287 have two different semantics: similarity for fuzzy sets associated with the instantiations of  
288 symbolic types and imprecision for those associated with the instantiations of numeric types.  
289 Consequently, we propose to realise those comparisons separately using two different  
290 measures: (i) a possibility degree of matching (noted  $\Pi$ ) and a necessity degree of matching  
291 (noted  $N$ ) which are classically used (see Dubois & Prade 1988) to compare a fuzzy set having  
292 a semantic of preference with a fuzzy set having a semantic of imprecision and (ii) an  
293 adequacy degree as proposed by (Baziz *et al.* 2006) to compare a fuzzy set having a semantic  
294 of preference with a fuzzy set having a semantic of similarity.

295 Let  $(a=v)$  be a selection criterion of the MIEL++ query  $Q$ ,  $v'$  a fuzzy annotation of the attribute  
296  $a$  (which is either a numeric type or a symbolic type of the ontology) stored in a RDF graph,  
297  $\text{sem}_{v'}$  the semantic of  $v'$ ,  $\mu_v$  and  $\mu_{v'}$  their respective membership degrees defined on the  
298 domain of values  $Dom$  (see definition in Table 1) associated with the attribute  $a$  and  $cl$  the  
299 function which corresponds to the fuzzy set closure. The comparison result depends on the  
300 semantic of the fuzzy set  $v'$ :

- 301 • if  $\text{sem}_{v'}=\text{imprecision}$ , the comparison result is given by the **possibility degree of**  
302 **matching** between  $v$  and  $v'$  noted  $\Pi(v,v')=\sup_{x \in Dom}(\min(\mu_v(x), \mu_{v'}(x)))$  and the  
303 **necessity degree of matching** between  $v$  and  $v'$  noted  $N(v,v')=\inf_{x \in Dom}(\max(\mu_v(x), 1 -$   
304  $\mu_{v'}(x)))$  (see Figure 5 for a graphical representation);

305 • if  $sem_v$ =similarity, the comparison result is given by the **adequacy degree** between  
306  $cl(v)$  and  $cl(v')$  noted  $ad(cl(v), cl(v')) = \sup_{x \in Dom} (\min(\mu_{cl(v)}(x), \mu_{cl(v')}(x)))$  (see Figure 6  
307 for a graphical representation).

308 The comparison results of fuzzy sets having the same semantic (similarity or imprecision) and  
309 associated with different selection criteria are aggregated using the min operator. Therefore,  
310 an answer to a query is a set of tuples composed of (i) the pertinence score  $ps$  associated with  
311 the queried semantic relation, (ii) three comparison scores associated with the selection  
312 criteria of the query: a global adequation score  $ad_g$  and two global matching scores  $\Pi_g$  and  $N_g$ ,  
313 and, (iii) the values associated with the answer attributes of the query. Based on those scores,  
314 we propose to define a total order on the answers which gives greater importance to the most  
315 pertinent answers compared with the ontology: answers are respectively sorted, in descendant  
316 order, according to  $ps$ ,  $ad_g$ ,  $N_g$  and  $\Pi_g$ .

317 Example: The answer to the MIEL++ query  $Q$  considered in the previous example and  
318 compared with the fuzzy annotations associated with the three rows of Table 2 is given below:

319  $\{ps=1, ad_g=0.5, N_g=1, \Pi_g =1, Microorg=(0.5/Clostridium Perfringens+0.5/Clostridium$   
320  $Botulinum), aw=[0.943, 0.95, 0.96, 0.97]\}$ ,

321  $\{ps=1, ad_g=0.5, N_g =0.5, \Pi_g=0.68, Microorg=(0.5/Staphylococcus spp.+0.5/Staphylococcus$   
322  $aureus), aw=[0.88, 0.98, 0.98, 0.99]\}$ ,

323  $\{ps=1, ad_g=0.5, N_g =0, \Pi_g=0.965, Microorg=(1.0/Salmonella), aw=[0.94, 0.99, 0.99, 0.991]\}$

324

### 325 **Application in bacterial growth simulations**

326 Microbial growth kinetics are usually described using primary models with four main  
327 parameters:  $x_0$  is the initial bacterial concentration,  $x_{max}$  is the maximum bacterial  
328 concentration, lag is the lag time (h), and  $\mu_{max}$  is the maximum specific growth rate ( $h^{-1}$ ). The

329 two last parameters vary according to the physico-chemical food characteristics and the  
 330 specific effect of the food matrix, whereas  $x_0$  and  $x_{\max}$  are considered as constant. The effects  
 331 (pH,  $a_w$ , storage temperature and food matrix) on  $\mu_{\max}$  is described by a multiplicative  
 332 function with interactions (Augustin *et al.*, 2005; Le Marc *et al.*, 2002) derived from the  
 333 cardinal models of Rosso *et al.* (1995):

334

$$335 \mu_{\max}(T, pH, a_w) = \mu_{opt} \cdot CM_2(T) CM_1(pH) CM_1(a_w) \xi(T, pH, a_w)$$

336 where

337

$$338 CM_n(X) = \begin{cases} 0 & , X \leq X_{\min} \\ \frac{(X - X_{\max})(X - X_{\min})^n}{(X_{opt} - X_{\min})^{n-1} [(X_{opt} - X_{\min})(X - X_{opt}) - (X_{opt} - X_{\max})(X_{opt} + X_{\min} - nX)]} & , X_{\min} < X < X_{\max} \\ 0 & , X \geq X_{\max} \end{cases}$$

$$339 \text{ and } \xi(T, pH, a_w) = \begin{cases} 1 & , \phi \leq 0.5 \\ 2 \cdot (1 - \phi) & , 0.5 < \phi < 1 \\ 0 & , \phi \geq 1 \end{cases} \quad \text{with} \quad \phi = \sum_i \frac{\omega(X_i)}{2 \cdot \prod_{j \neq i} (1 - \omega(X_j))} \quad \text{and}$$

$$340 \omega(X) = \left( \frac{X_{opt} - X}{X_{opt} - X_{\min}} \right)^3,$$

341

342 where  $\mu_{\max}$  is the specific growth rate in the considered food and in the considered  
 343 temperature, pH and  $a_w$  conditions,  $\mu_{opt}$  is the optimal growth rate value when temperature,  
 344 pH and  $a_w$  are set to their optimal values, and  $X_{\min}$ ,  $X_{opt}$  and  $X_{\max}$  are minimal, optimal and  
 345 maximal temperatures, pH, and water activities for growth.  $n$  is the shape parameter of the  
 346 CM model.

347 The lag time (lag) is calculated according to the following equation:

$$348 lag = \frac{\mu_{opt} \cdot lag_{\min}}{\mu_{\max}}$$



349 where  $\text{lag}_{\min}$  is the lag time value when temperature, pH and  $a_w$  are set to their optimal values.  
350 The optimal growth rate  $\mu_{\text{opt}}$  and the minimal lag time  $\text{lag}_{\min}$  depend on both strain and food  
351 matrix (Pinon *et al.*, 2004).  
352 Cardinal values are characteristic parameters of the microorganisms and are independent from  
353 the food matrix. Consequently,  $\text{CM}_n(X)$  functions are calculated using only bacterial species  
354 parameters and physico-chemical factors, whereas  $\mu_{\text{opt}}$  and  $\text{lag}_{\min}$  can be estimated from an  
355 experimental  $\mu_{\text{max}}$  and lag related to microbial growth in the food to be assessed:

$$356 \mu_{\text{opt}} = \frac{\mu_{\text{max}}(T, \text{pH}, a_w)}{\text{CM}_2(T) \cdot \text{CM}_1(\text{pH}) \cdot \text{CM}_1(a_w)}$$

$$357 \text{lag}_{\min} = \frac{\mu_{\text{max}} \cdot \text{lag}}{\mu_{\text{opt}}}$$

358  $\mu_{\text{opt}}$  and  $\text{lag}_{\min}$  calculations need at least one  $\mu_{\text{max}}$  and one lag value at certain temperature, pH  
359 and  $a_w$  conditions. In the Sym'Previous calculation tools, these data can be entered following  
360 two main ways. On the one hand, the user can enter manually values if he/she has this  
361 information (results of a challenge-test, bibliographic data, etc.). On the other hand, the user  
362 can select the food product in the ontology (figure 4), and submit a query simultaneously in  
363 local and web data. The results are returned in a synthetic table (figure 7) where experimental  
364  $\mu_{\text{max}}$  and lag values are proposed to be included in the  $\mu_{\text{opt}}$  and  $\text{lag}_{\min}$  calculation.

365 In the worst case (no food data available), simulations are carried out with known  $\mu_{\text{opt}}$  and  
366  $\text{lag}_{\min}$  obtained in culture media in laboratory conditions.

367

## 368 **Results and discussion**

369 We use in this paper the same corpus as the one used in Hignette *et al.* (2008) which  
370 composed of 60 tables extracted from publications on food microbiology, in order to test our

371 instantiation of semantic relations in Web data tables and the flexible querying of the Web  
372 data tables such annotated.

373 We have automatically instantiated the 119 relations which were correctly recognised to  
374 annotate the 60 tables in the experiment presented in Hignette *et al.* (2008).

375

#### 376 *Instantiation of numeric types*

377 The instantiation of numeric types was analysed for the first data row of each table: we  
378 assume that the structure is enough homogenous inside a table, such that the instantiation of  
379 the first row can be considered as representative of what happens in the whole table. On the  
380 119 relations, there were two errors on the extraction of numeric values (one was an error of  
381 type recognition; one was an error of numeric value recognition). For 5 tables (corresponding  
382 to 13 relations), the numeric type *Temperature* was not instantiated because its value was not  
383 present in the table but in the textual environment of the table in the original publication.  
384 There also were three errors in interval reconstruction (values were considered as isolated  
385 while it was an interval) and one error in the construction of a minimum/optimum/maximum  
386 trapezoid fuzzy set (values were considered as isolated). For all 100 remaining relations, all  
387 numeric values were correctly instantiated.

388

#### 389 *Instantiation of symbolic types*

390 The quality of the instantiation of symbolic types was evaluated on 185 instances of food  
391 products extracted from the corpus of 60 tables. For those food products, the “best match” in  
392 the ontology (i.e. the term in the ontology that was the nearest to the meaning of the term in  
393 the table) was manually defined. The evaluation is done by looking at the position of the “best  
394 match” in the automatic instantiation, by order of descendant membership degree. The  
395 position is evaluated at worse, i.e. if there are several terms in the ontology having the same

396 membership degree in the fuzzy set used for the instantiation, the “best match” is always  
397 considered as being in last position. This evaluation at worse comes from the need to  
398 manually validate the instantiations: if we ask a user to choose among the 5 terms having the  
399 best membership degree, we want to be sure that the “best match” is among those 5. On the  
400 185 terms from the web, 78% had a “best match” for which the term similarity with the term  
401 from the web was not null, 46% had their “best match” in first position in the computed  
402 instantiation, while 66% had their “best match” among the five best positions. These results  
403 validate the approach of keeping a fuzzy set for instantiating the symbolic types, instead of  
404 keeping only the term in the taxonomy having the best similarity with the term in the table.

405

#### 406 *Flexible querying of the RDF database including Web data tables*

407 In preliminary tests performed on the RDF graph database composed of more than 22.000  
408 RDF triples (312 graphs), we have evaluated 5 queries (see Table 3) covering at least 50% of  
409 the database entries. Querying quality is assessed using two measures: precision and recall.  
410 Precision is the ratio of correct answers over the total number of computed answers. Recall is  
411 the ratio of correct computed answers over the number of expected answers. We obtain better  
412 results in the queries where the selection criterion concerns microorganisms than in the ones  
413 concerning food products. This is due to the fact that microorganism names are more  
414 standardized in Web data tables than food product names. Nevertheless, we obtain a precision  
415 of 100% for the two last queries concerning food product if we put a threshold of 0.7 on the  
416 terms similarity degrees.

417

#### 418 *Microbial growth simulation in food*

419 Growth simulation in food requires data related to the bacterial species (cardinal values),  
420 physico-chemical properties of food (pH, water activity, storage temperature) and food matrix

421 effect. The implementation of the food matrix in models is achieved thanks to the  $\mu_{opt}$  and  
422  $lag_{min}$  parameters. These two parameters can be estimated only from experimental data  
423 collected in food. In the absence of food experimental data, simulations are usually processed  
424 using  $\mu_{opt}$  and  $lag_{min}$  obtained in culture media at laboratory conditions, giving approximate  
425 results.

426 The following example presents *Listeria monocytogenes* growth simulation in cold smoked  
427 salmon (pH 6 +/- 0.2 and water activity 0.970 +/- 0.003) and stored at 6°C (+/- 1°C),  
428 performed with the Sym'Previus probabilistic software (Couvert *et al.*, submitted article). The  
429 initial contamination for simulation is expected to -2 log CFU/g (+/- 0.2). On the one hand,  
430 without any experimental data related to *L. monocytogenes* growth in cold smoked salmon,  
431 the simulation includes physico-chemical properties and optimum growth parameters obtained  
432 in culture media (figure 8-A). On the other hand, a query in local and web data allows one to  
433 retrieve experimental data (figure 7) which are used to estimate  $\mu_{opt}$  and  $lag_{min}$  and,  
434 consequently, taking into account the food matrix in simulations. The figure 8-B reproduces  
435 the previous simulation in cold smoked salmon, taking into account the food product  
436 parameters. Median population size after 21 days of storage reaches 4.4 log CFU/g in  
437 laboratory media simulations, and -0.09 log CFU/g taking into account bibliographic data in  
438 cold smoked salmon. These results demonstrate the importance of the food matrix effect on  
439 bacterial growth, and the necessity to link bibliographic data stored in databases with  
440 simulation softwares.

441

## 442 *Discussion*

443 Recent propositions in the Semantic Web community propose to extract, filter, annotate and  
444 query Web data tables (Ying *et al.* 2007, Cafarella *et al.*, 2008), but they have not been  
445 designed with the same objectives. TableSeer (Ying *et al.* 2007) for instance permits to extract

446 a set of predefined metadata (caption, cell content, geographical position of the table in the  
447 HTML page, ...) from Web tables, but it does not compare the schema of the Web tables to  
448 preexisting schemas defined in an ontology. We can also cite WebTables (Cafarella *et al.*,  
449 2008) which proposes a system to identify data tables in the huge amount of tables included in  
450 HTML documents and index them, in order to query and rank them. Nevertheless, the  
451 WebTables querying language is only composed of a set of key-words which are compared  
452 with the titles of the columns present in the Web tables. The row content of the Web tables is  
453 not used in the querying process which is only based on global co-occurrences frequencies  
454 statistics of the terms appearing in the titles of the columns. Therefore, it is not possible to  
455 compare the results we obtain with our annotation and instantiation methods with other Web  
456 table mining methods because they don't have the same aim.

457

## 458 **Conclusion and perspectives**

459 Probabilistic simulations of Sym'Previus software needs a lot of data in food products to take  
460 the food matrix into account and to assess food variability in bacterial growth simulations. A  
461 prototype of the @WEB and the MIEL++ tools will be soon integrated with the predictive  
462 modeling tools of the Sym'Previus project. These automatic links between web data and  
463 simulation tools are a major contribution to enhance risk assessment. In the near future, we  
464 will study ontology evolution in order to take into account the evolutions of the predictive  
465 modeling tool. For example, (i) the impact of the packaging and gaz transfer on the behaviour  
466 of the microorganism in the food matrix should be studied or (ii) the possible use of the  
467 growing number of ontologies which are available on the Web in close domains ([OBOE](#),  
468 [OUM](#), ...) and which may enhance the quality of our annotation and instantiation methods if  
469 they are merged with our own ontology. Therefore, in order to take into account this new  
470 information, the domain ontology of our system should evolve, and, new methodologies and

471 tools should also be developed according to this ontology evolution. Another perspective is to  
472 extend this work in order to also be able to extract pertinent information represented as  
473 graphics or directly from the text.

474

## 475 **Acknowledgements**

476 Financial support from the French National Research Agency (ANR) for the project  
477 WebContent in the framework of the National Network for Software Technology (RNNTL) is  
478 gratefully acknowledged.

479

## 480 **References**

481 Augustin, J. C., Zuliani, V., Cornu, M., Guillier, L., 2005. Growth rate and growth probability  
482 of *Listeria monocytogenes* in dairy, meat and seafood products in suboptimal conditions.  
483 Journal of Applied Microbiology. 99, 1019-1042.

484 Baranyi, J., Tamplin, M., 2004. ComBase: A Common Database on Microbial Responses to  
485 Food Environments. Journal of Food Protection. 67, 1834-1840.

486 Baziz, M., Boughanem, M., Prade, H., Pasi, G., 2006. A fuzzy logic approach to information  
487 retrieval using a ontology-based representation of documents. In Sanchez, E. (Ed), Fuzzy logic  
488 and the Semantic Web, Elsevier, pp. 363–377.

489 Buche, P., Dervin, C., Haemmerlé, O., Thomopoulos, R., 2005. Fuzzy querying of  
490 incomplete, imprecise, and heterogeneously structured data in the relational model using  
491 ontologies and rules. IEEE Transactions on Fuzzy Systems. 13, 373-383.

492 Buche, P., Dibia-Barthélemy, J., Haemmerlé, O., Hignette, G., 2006. Fuzzy semantic tagging  
493 and flexible querying of XML documents extracted from the Web. Journal of Intelligent  
494 Information System. 26, 25-40.

495 Cafarella, M. J., Halevy, A. Y., Zhe Wang D., Wu, E., Zhang, Y., 2008. WebTables:  
496 exploring the power of tables on the web. Proceedings of the VLDB Endowment Vol 1, Issue  
497 1 538-549.

498 Couvert, O., Augustin, J. C., Buche, P., Carlin, F., Coroller, L., Denis, C., Jamet, E., Mettler,  
499 E., Pinon, A., Stahl, V., Zuliani, V., Thuault, D., 2007. Optimising food process and  
500 formulation through Sym'Previus, managing of the food safety. Proceedings of 5<sup>th</sup>  
501 International Conference Predictive Modeling in Foods.

502 Dubois, D., Prade, H., 1988. In: Possibility theory- An approach to computerized processing  
503 of uncertainty. Plenum Press, New York.

504 Dubois, D., Prade, H., 1997. The three semantics of fuzzy sets. Fuzzy Sets and Systems.  
505 90(2), 141-150.

506 Hignette, G., Buche, P., Couvert, O., Dibie-Barthélemy J., Doussot, D., Haemmerlé, O.,  
507 Mettler, E., Soler, L., 2008. Semantic annotation of Web data applied to risk in food.  
508 International Journal of Food Microbiology. 128, 174-180.

509 Le Marc, Y., Huchet, V., Bourgeois, C. M., Guyonnet, J. P., Mafart, P., Thuault, D., 2002.  
510 Modeling the growth kinetics of *Listeria* as a function of temperature, pH and organic acid  
511 concentration. International Journal of Food Microbiology. 73, 219-237.

512 Lin, D., 1998. An information-theoretic definition of similarity. In: ICML '98 : Proceedings  
513 of the Fifteenth International Conference on Machine Learning, 296-304. Morgan Kaufmann  
514 Publishers Inc., San Francisco.

515 McMeekin, T. A., Baranyi, J., Bowman, J., Dalgaard, P., Kirk, M., Ross, T., Schmid, S.,  
516 Pinon, A., Zwietering, M. H., Perrier, L., Membré, J. M., Leporq, B., Mettler, E., Thuault, D.,  
517 Coroller, L., Stahl, V., Vialette, M., 2004. Development and Validation of Experimental  
518 Protocols for Use of Cardinal Models for Prediction of Microorganism Growth in Food  
519 Products. Applied and Environmental Microbiology. 70, 1081-1087.

520 Rosso, L., Lobry, J. R., Bajard, S., Flandrois, J. P., 1995. Convenient Model To Describe the  
521 Combined Effects of Temperature and pH on Microbial Growth. *Applied and Environmental*  
522 *Microbiology*. 61, 610-616.

523 Tamplin, M., Baranyi, J., Paoli, G., 2003. Software programs to increase the utility of  
524 predictive microbiology information. In: McKellar, R.C., Lu, X. (Eds), *Modeling Microbial*  
525 *responses in Foods*. CRC, Boca Raton.

526 Thomopoulos R., Buche P., Haemmerlé O., 2003. Different kinds of comparisons between  
527 fuzzy conceptual graphs. *Lecture Notes In Artificial Intelligence*. 2746, 54-68.

528 Thomopoulos R., Buche P., Haemmerlé O., 2006. Fuzzy sets defined on a hierarchical  
529 domain. *IEEE Transactions on Knowledge and Data Engineering*. 18(10), 1397-1410.

530 Ying, L., Kun, B., Prasenjit, M., Lee Giles, C., 2007. TableSeer: automatic table metadata  
531 extraction and searching in digital libraries. *Proceedings of the 7th ACM/IEEE-CS joint*  
532 *conference on Digital libraries*, 91-100 (ISBN:978-1-59593-644-8).

533 Zadeh, L., 1965. Fuzzy sets. *Information and control*. 8, 338-353.

534 Zadeh, L., 1978. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*. 1, 3-  
535 28



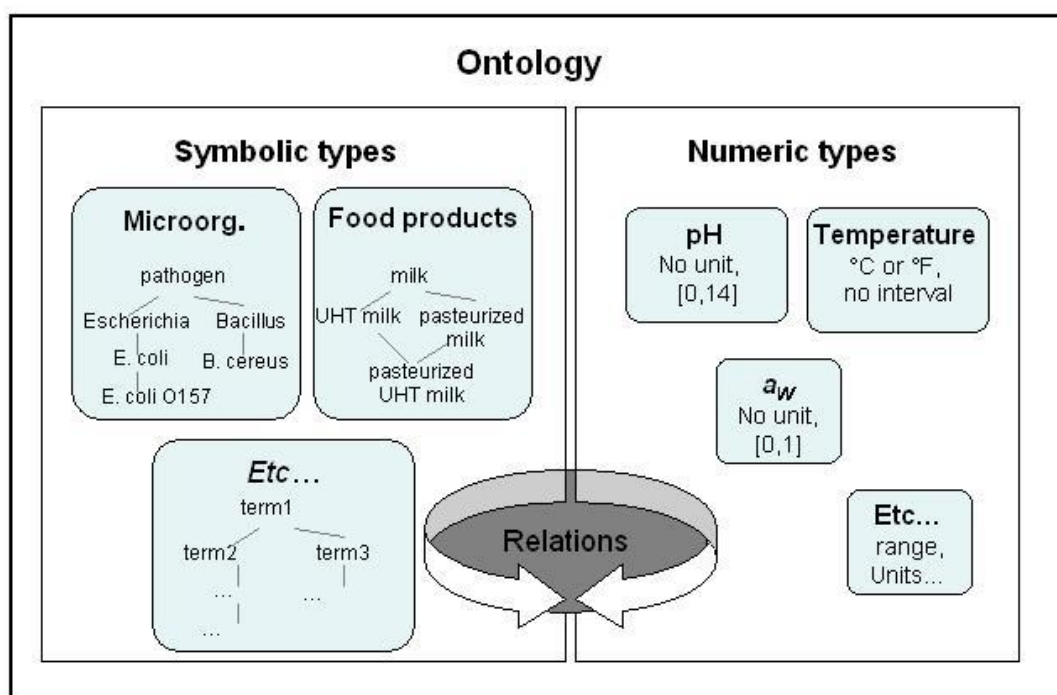


Figure 1

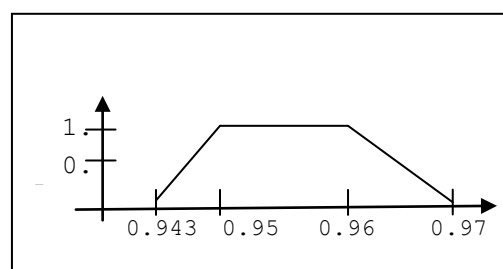
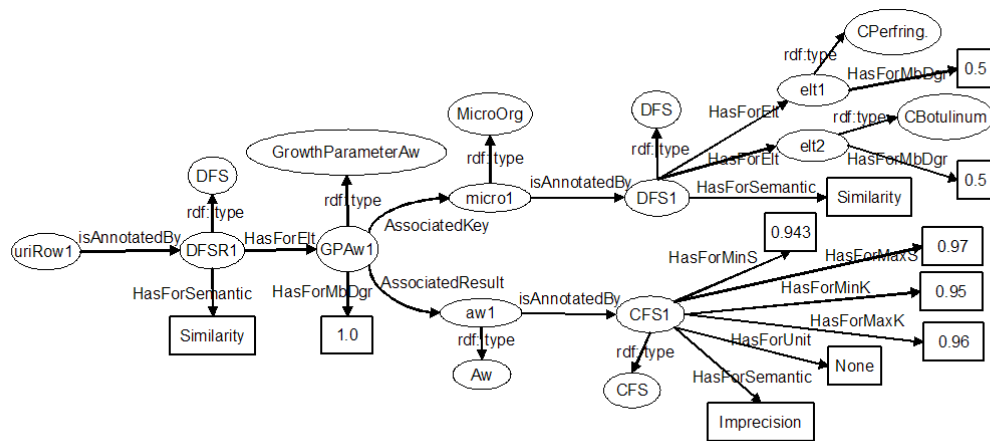


Figure 2

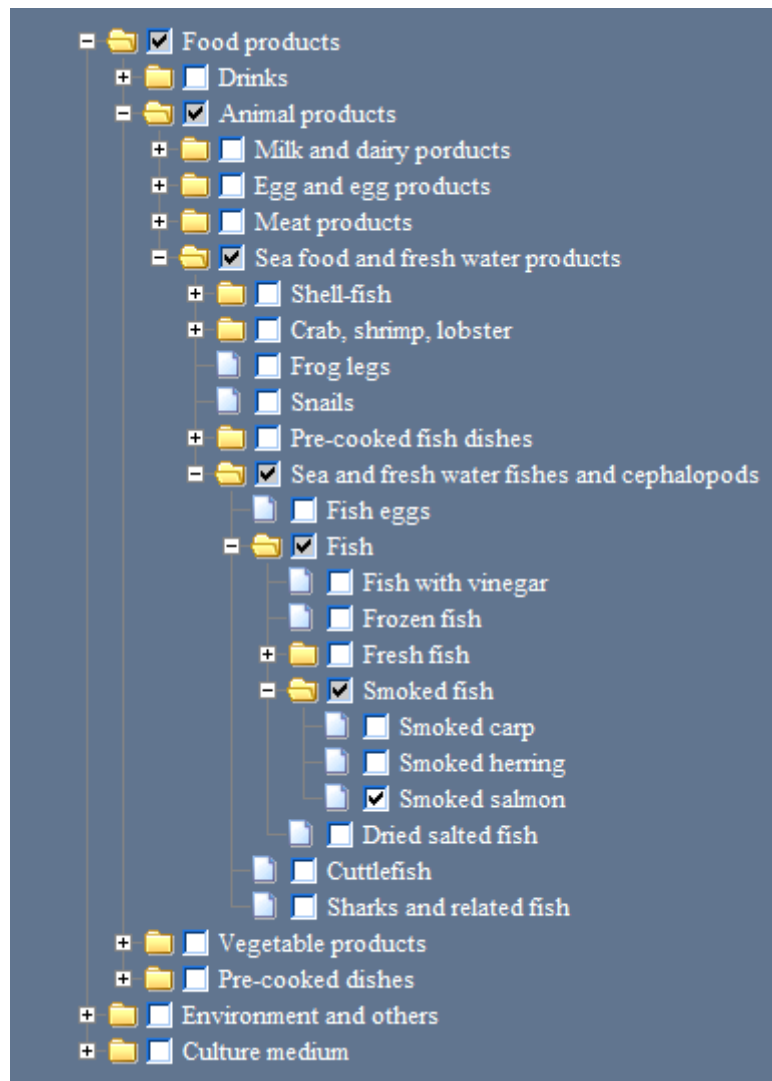


547

548

Figure 3

549

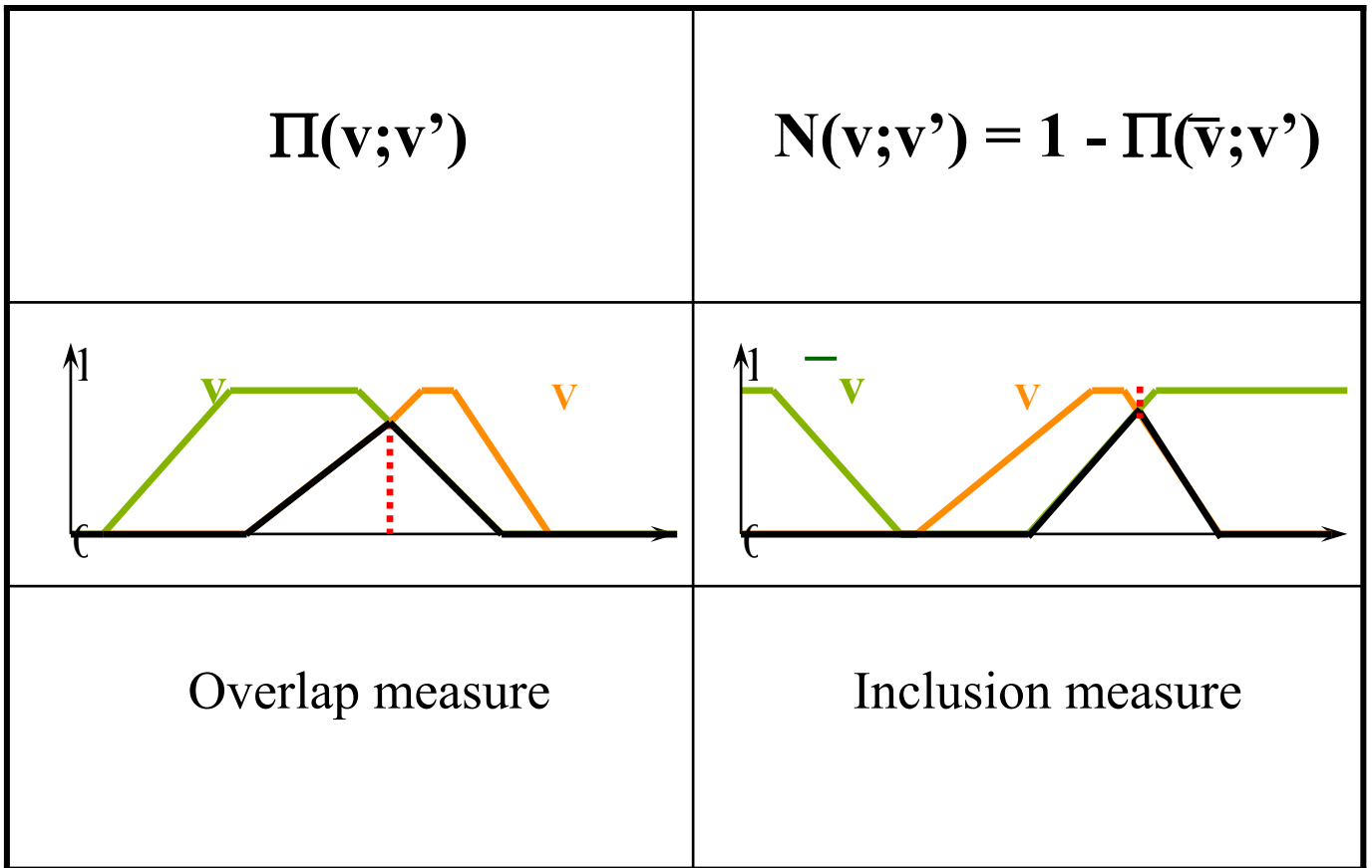


550

551

Figure 4

Figure 5

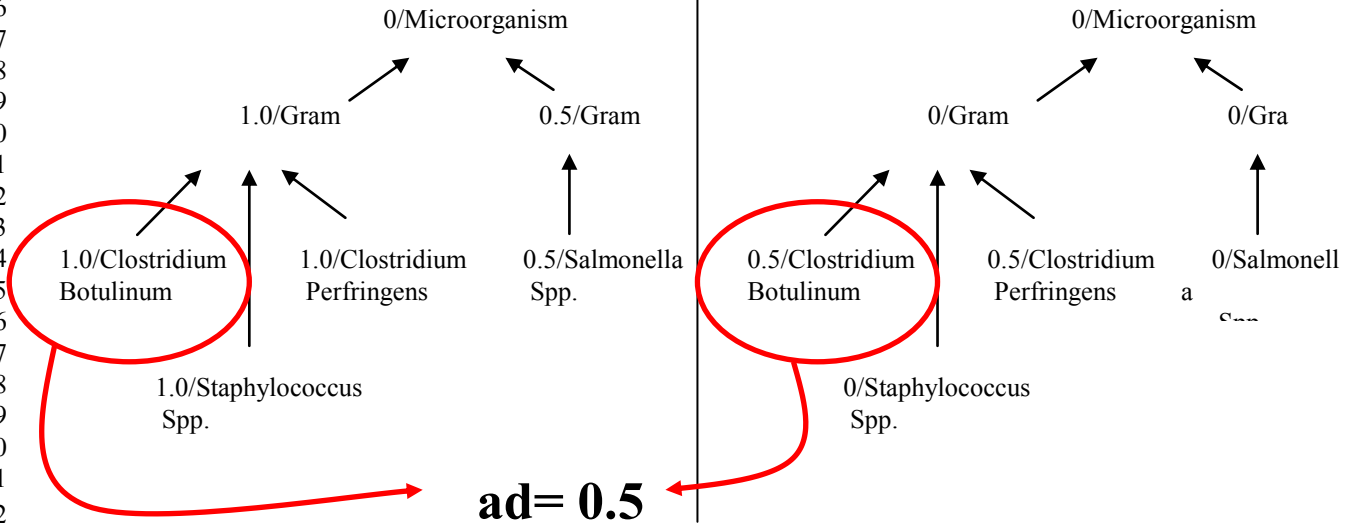


Version postprint

Figure 6

### clos(MicroPreferences)

### clos(DFSR1)



#	Food	Factors				Remarks and source	Parameters				kinetic	Accuracy	Include
		T°	pH	aw	Others		lag	$\mu_{max}$	No	Nmax			
1	Saumon fumé	.	6.1	<input type="checkbox"/>	<a href="#">Click here</a>	<a href="#">Click here</a>	127.8	.	.	.	.	.	<input type="checkbox"/>
2	Saumon fumé	.	6.1	<input type="checkbox"/>	<a href="#">Click here</a>	<a href="#">Click here</a>	25.33	.	.	.	.	.	<input type="checkbox"/>
3	Saumon fumé	.	6.1	<input type="checkbox"/>	<a href="#">Click here</a>	<a href="#">Click here</a>	51.3	.	.	.	.	.	<input type="checkbox"/>
4	Saumon fumé	.	6.1	<input type="checkbox"/>	<a href="#">Click here</a>	<a href="#">Click here</a>	317.7	.	.	.	.	.	<input type="checkbox"/>
5	Saumon fumé	.	6.1	<input type="checkbox"/>	<a href="#">Click here</a>	<a href="#">Click here</a>	.	0.103	.	.	.	.	<input type="checkbox"/>
6	Saumon fumé	.	6.1	<input type="checkbox"/>	<a href="#">Click here</a>	<a href="#">Click here</a>	.	0.086	.	.	.	.	<input type="checkbox"/>
7	Saumon fumé	.	6.1	<input type="checkbox"/>	<a href="#">Click here</a>	<a href="#">Click here</a>	.	0.033	.	.	.	.	<input type="checkbox"/>
8	Saumon fumé	.	6.1	<input type="checkbox"/>	<a href="#">Click here</a>	<a href="#">Click here</a>	.	0.036	.	.	.	.	<input type="checkbox"/>
9	Saumon fumé	10.0	6.9	0.989	.	<a href="#">Click here</a>	25.5	0.0178	4.33	36.3	Ok	+	<input type="checkbox"/>
10	Saumon fumé	15.0	6.9	0.989	.	<a href="#">Click here</a>	27.1	0.0716	4.32	8.31	Ok	++	<input type="checkbox"/>
11	Saumon fumé	25.0	6.9	0.989	.	<a href="#">Click here</a>	8.42	0.231	4.45	8.59	Ok	++	<input type="checkbox"/>
12	Saumon fumé	10.0	6.9	0.989	.	<a href="#">Click here</a>	.	.	.	.	Fit	.	<input type="checkbox"/>
13	Saumon fumé	15.0	6.9	0.989	.	<a href="#">Click here</a>	.	.	.	.	Fit	.	<input type="checkbox"/>
14	Saumon	25.0	6.9	0.989	.	<a href="#">Click here</a>	.	.	.	.	Fit	.	<input type="checkbox"/>

603

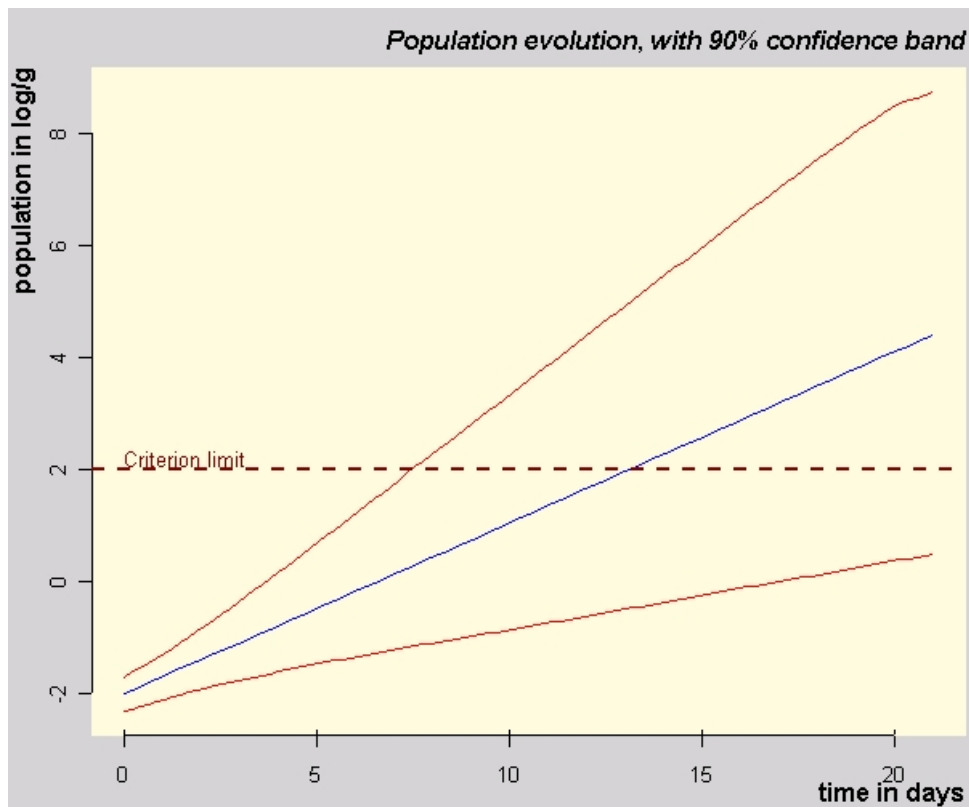
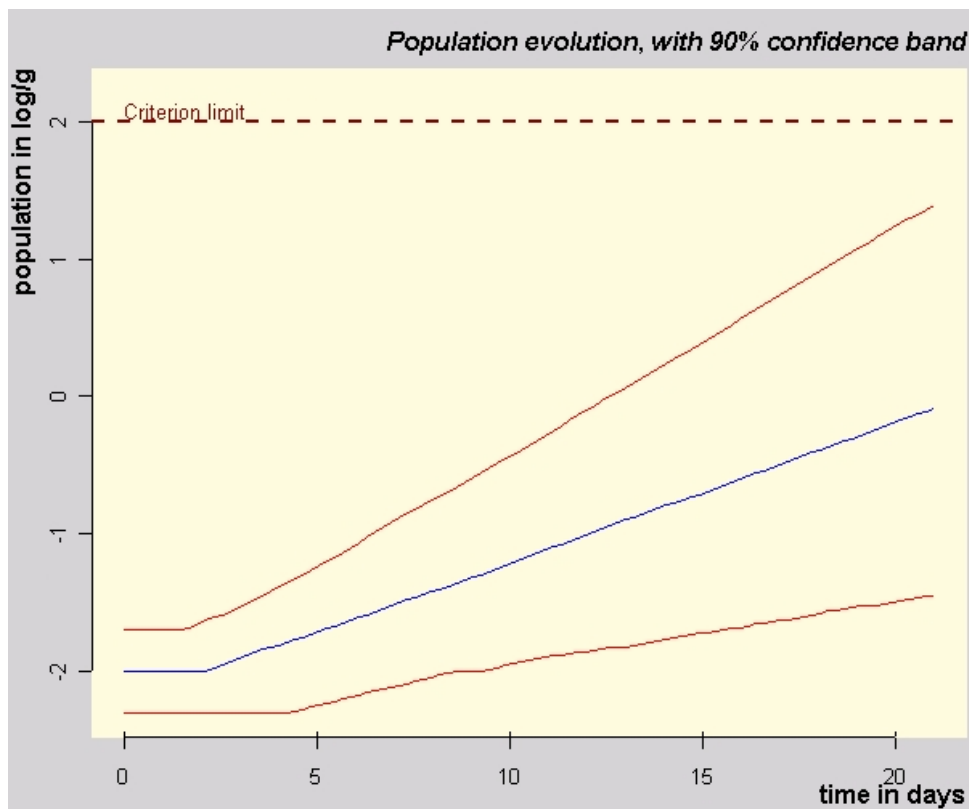
604

605

606

607

Figure 7

608  
609610  
611  
612

613

Figure 8

614

615 Table 1: Glossary of technical terms

Term	Explanation
Ontology	It is, for a given scientific domain, a set of concepts and semantic relations which link those concepts. By example, <i>Microorganism</i> and <i>Clostridium Perfringens</i> , <i>pH</i> , <i>a<sub>w</sub></i> are concepts of the ontology. <i>Microorganism</i> is a concept classified as symbolic data type. <i>pH</i> is a concept classified as numeric data type. <i>Microorganism</i> and <i>Clostridium Perfringens</i> are linked by the <i>a kind of</i> semantic relation.
Semantic relation	It is a relation which links concepts of the ontology. Semantic relations are defined by their signature which is composed of a result data type and a set of access data types. For example, the relation <i>GrowthParameterAw</i> , representing the growth limits of a microorganism for any food product, has for access type the symbolic type <i>Microorganism</i> and for result type the numeric type <i>a<sub>w</sub></i> .
Instanciation	The instanciation of a concept or a semantic relation is an occurrence of a concept (a numeric or a symbolic type) or a semantic relation used to annotate a given row of a given Web data table.
Domain of values (Dom)	A domain of values is defined for a symbolic type and a numeric type of the ontology. The domain of values of a symbolic type is its taxonomy of possible values in the ontology. The domain of values of a numeric type is its numeric range in the ontology

616

617 Table 2: Cardinal values (growth boundaries).

Organism	<i>a<sub>w</sub></i> minimum	<i>a<sub>w</sub></i> optimum	<i>a<sub>w</sub></i> maximum
<i>Clostridium</i>	0.943	0.95-0.96	0.97
<i>Staphylococcus</i>	0.88	0.98	0.99
<i>Salmonella</i>	0.94	0.99	0.991

618

619

620 Table 3: Evaluation of query results

Queried relation	Selection criteria	Precision-recall	Nb of answer graphs
Lag Time	Microorganism= <i>L. monocytogenes</i>	100%-100%	47 graphs
Lag Time	Microorganism= <i>P. fluorescens</i>	100%-100%	29 graphs
Growth kinetics	Microorganism= <i>E. coli</i>	100%-100%	39 graphs
Lag Time	FoodProduct= Egg salad	50%-100%	24 graphs
Growth kinetics	FoodProduct= Salad	54%-100%	26 graphs

621

622