



HAL
open science

Combining fuzzy querying of imprecise data and predictive microbiology using category-based reasoning for prediction of the possible microbial spoilage in foods: application to [*i*]Listeria monocytogen[/*i*]es

Patrice Buche, Catherine Dervin, Agnès Brouillaud-Delattre, Nathalie Gnanou-Besse

► **To cite this version:**

Patrice Buche, Catherine Dervin, Agnès Brouillaud-Delattre, Nathalie Gnanou-Besse. Combining fuzzy querying of imprecise data and predictive microbiology using category-based reasoning for prediction of the possible microbial spoilage in foods: application to [*i*]Listeria monocytogen[/*i*]es. International Journal of Food Microbiology, 2002, 73 (2-3), pp.171-185. 10.1016/S0168-1605(01)00647-X . hal-01229554

HAL Id: hal-01229554

<https://hal.science/hal-01229554>

Submitted on 16 Nov 2015

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 **Combining fuzzy querying of imprecise data and predictive microbiology**
2 **using category-based reasoning**
3 **for prediction of the possible microbial spoilage in foods :**
4 **application to *Listeria monocytogenes*.**

5
6 **PATRICE BUCHE*¹, CATHERINE DERVIN¹,**
7 **AGNES BROUILLAUD-DELATTRE³, NATHALIE GNANOU-BESSE²,**

8 ¹UMR INA P-G/INRA, BIA Paris
9 16, rue Claude Bernard 75231 Paris Cédex 5 FRANCE
10 Tel 331 44 08 16 75, Fax 331 44 08 16 66
11 E-mail: Patrice.Buche@inapg.inra.fr

12
13 ²Agence Française de Sécurité Sanitaire des Aliments, Laboratoire d'Etudes et de Recherches
14 sur l'Hygiène et la Qualité des Aliments, 39/41, rue du 11 novembre 94700 Maisons-Alfort
15 ³Agence Française de Sécurité Sanitaire des Aliments/Ecole Nationale Vétérinaire de
16 Toulouse , 23 chemin des Capelles 31076 Toulouse cedex 3

17
18 * corresponding author

19 **Abstract**

20 Various predictive models of microbial behavior have been created and extensive data
21 collection has been done by numerous private or public laboratories. However, significant
22 differences between predicted and observed values in foods have been observed and need to
23 be stressed, understood and explained as much as possible. In this paper, we present a
24 software tool (currently at the level of a prototype) able : (i) to store in a database all relevant
25 information expressed on one hand as qualitative or quantitative data and on the other hand as
26 precise or imprecise data, (ii) to retrieve the more relevant information from the database
27 using queries where criteria may be expressed as fuzzy values in order to enhance the
28 flexibility of the search, (iii) to compute, in addition to the nearest data, an estimation of
29 searched values using statistical models. The architecture of this software tool is structured as

30 a category-based reasoning system. Example queries about *Listeria monocytogenes* (*L.*
31 *monocytogenes*) illustrate functionalities of this tool.

32 **Key words :** Predictive microbiology, *Listeria monocytogenes*, Fuzzy logic, Information
33 retrieval system

34 **Introduction**

35 Numerous factors affect microbial growth in food systems and often synergistic or
36 antagonist interactions are significant. Thus, microbial growth in foods can be enhanced or
37 inhibited depending on multiple food ingredients and storage conditions (Eifert et al., 1996).
38 This can explain why an understanding of microbial behavior usually requires large-scale
39 experiments under a wide range of conditions and involving extensive data collection.

40 Since data collection can be very time-consuming and labor intensive, numerous
41 laboratories have developed predictive growth models, especially in Australia, the U.K., the
42 U.S.A. and France with particular reference to *L. monocytogenes*.

43 *L. monocytogenes* has been associated in the last 15 years with several food-borne illness
44 outbreaks involving a large range of foods such as raw milk, cheeses (particularly soft-ripened
45 varieties), raw vegetables, fermented raw-meat sausages, raw or cooked meats. Microbial
46 safety of many food products depends on refrigeration after commercialization, as most
47 microorganisms do not grow at refrigeration temperatures. However this is not the case for *L.*
48 *monocytogenes* because of its ability to grow at temperatures as low as 3°C (even 0.1°C and –
49 0.4°C (Bajard et al., 1996).

50 In order to evaluate food safety relating to the *Listeria* growth hazard in "chilled prepared
51 food products", temperature, pH and water activity are the main factors that must be taken into
52 account in predictive models (Bajard et al., 1996). Various models are able to predict lag
53 times and growth rates of *L. monocytogenes* (or other microorganisms) in many food matrices

54 at constant temperatures (Ratkowsky et al., 1983; Zwietering et al., 1991; Duh and Schaffner,
55 1993; Rosso et al., 1993; Baranyi et al., 1993) and a small number of models also do so in a
56 dynamic temperature environment (Baranyi et al., 1995; Van Impe et al., 1995; Alavi et
57 al.,1999). Most of them have improved their predictions by including other environmental
58 factors such as the presence of organic acids, the nature of packaging atmosphere (Buchanan
59 and Phillips, 1990; Murphy et al., 1996; Fernandez et al., 1997) or some other biological
60 factors such as the physiological state of inoculum cells and inoculum size (Mc Kellar et al.,
61 1997).

62 However, some authors have observed significant differences between predicted values and
63 observed values in foods and suggest that these differences are due to factors not taken into
64 account in existing models (Brouillaud-Delattre et al., 1997; Dalgaard and Jorgensen, 1998).
65 For example, the growth response of *L. monocytogenes* can be modified by the strain
66 considered (Siswanto and Richard, 1992; Begot et al., 1997), the accumulation of inhibiting
67 substances (Guinot-Thomas et al., 1995), initial contamination level (Gay et al. 1996),
68 competitive interactions (Farrag and Marth, 1989) or microstructural texture (Guerzoni et al.,
69 1994; Brouillaud-Delattre et al., 1997).

70 In some cases, failure of predictive models is due to insufficient knowledge of bacterial
71 behavior in complex environments such as food matrices. However, particularly over the last
72 few years, a lot of studies have provided better knowledge of new factors that may
73 significantly influence the growth of bacteria. The main problem concerns the quality of these
74 data : they might be imprecise and incomplete. On the one hand, experimental results are
75 often synthesized as intervals of values obtained with or without statistical methods, in order
76 to represent variability, then the data are imprecise. On the other hand, experimental results
77 exist only for a limited amount of substrates, then the data are incomplete, in comparison to
78 the set of substrates on which users may query. The use of an computerized logical reasoning

79 may help to manage this incomplete and potentially imprecise information (Dubois and Prade,
80 1995).

81 The object of this study is to develop a "logical tool" to include the incomplete and
82 imprecise information provided by the literature in the database of a "decision support system
83 for prediction of microbial risk". Previous works have already proposed such a kind of system.
84 Zwietering et al., (1992) takes only into account the fact that information is incomplete for
85 food products. He does not propose any solution to represent and manage imprecise
86 information. Estimation is only proposed for growth phenomena using a universal model.
87 Adair and Briggs (1993) presents only the concept and main stages of a dedicated expert
88 system.

89 Our logical approach suggests a new kind of predictive response. First, we propose a
90 unified framework based on fuzzy logic and theory of possibility which permits to : (i) express
91 the user's query preferences (and therefore to control the possible enlargement process of
92 database scanning) (ii) represent and manage the imprecise information. Second, the system
93 presents the ability of delivering several hierarchical responses (experimental results obtained
94 from bibliography, unpublished experimental results, ...) brought by the database with an
95 indication of the respective degree of matching for each response. Third, the system is able to
96 propose a fuzzified estimated response using predictive models stored in its knowledge base.
97 Fourth, a case based reasoning system identifies the category of queried data. This support
98 system is part of a national project on predictive microbiology in food systems.

99 The originality of our approach is demonstrated using as an example to *L. monocytogenes*
100 in this paper. But, our method and our tool are completely independent to any particular
101 bacterium and can be used whatever studied bacterium.

102 **Materials and methods**

103 In this paper, we are concerned on one hand with the storage and handling of imprecise
104 information and on the other hand with the expression of querying using fuzzy values to
105 retrieve the more relevant information available in the database. In the first section, we
106 introduce the relational model chosen to structure the information in the database. In the
107 second section, we present the unique framework selected to represent on one hand query
108 preferences and on the other hand precise or imprecise values. In the third section, we present
109 statistical models we use to compute estimated values delivered to the user in addition to the
110 closest information found in the database. In the fourth section, we introduce resemblance
111 relations used to retrieve the closest information stored in the database in addition to query
112 preferences. Finally, in the fifth section, we present the main concepts of the information
113 retrieval system we have designed.

114 **Information structured using the relational model.**

115 The relational model is the most popular model used all over the world to represent
116 information in databases. It was proposed by Codd (1970) in the early 70's and it has been
117 available as a commercial product for 20 years. It manages large quantities of data using a
118 DBMS - Database Management System- such as Oracle for instance as well as personal
119 databases using a DBMS such as Microsoft Access. The basic concept is very simple :
120 elementary information is gathered in *relations* which represent *objects* of the modeled world
121 or *links* between these objects. The concept of relation is implemented as a table. For instance,
122 a bacterium is an object of the world we want to model. Elementary information such as the
123 name of the bacterium, the name of the strain, its identification (unique) number in the
124 database are gathered in a table called *Bacterium* and represent the object Bacterium in the
125 database. These 3 pieces of information are called attributes of the table. A domain of values
126 is associated to each attribute. For instance, the attribute *BacteriumName* must be a string as

127 the attribute *BacteriumId* must be a positive integer. All the bacteria represented in the
128 database are described by the values associated to these 3 attributes and are called tuples of
129 the table *Bacterium*. Now, if we want to represent the fact that a bacterium produces a
130 bacteriocin, we will add two new tables. The table called *Bacteriocin* will represent the object
131 Bacteriocin. Another table will represent the semantic link between the bacterium and the
132 bacteriocin ; this table will be called *IsProducedBy*.

133 The relational model is built on solid theoretical foundations and useful normalization rules
134 have been provided to help the administrator of the database to define the table structures.
135 But, an important drawback of the relational model is that the modeled world is split into a
136 great number of tables and the database structure composed of a long list of tables is very
137 complex to understand. This is the reason why the concept of view has been introduced. A
138 view is a *virtual* table in which all the information needed by the user is brought together. For
139 instance, we can define the view *BacteriocinProducedByBacterium* which brings together the
140 information split into the 3 tables *Bacterium*, *Bacteriocin* and *IsProducedBy*. Only the
141 administrator of the database needs to know the real and entire structure of the database. A
142 user only knows the view designed by the administrator to access the data he needs.

143 **Query preferences and imprecision representation.**

144 The fuzzy logic theory base approach presented in Zadeh (1965) has been largely used to
145 represent values associated to selection attributes in order to enhance query flexibility (Cox,
146 1995; Bosc et al., 1999). More precisely, fuzzy sets are mathematical objects here used with a
147 semantic of preferences. For example, if one queries a database with the selection attribute
148 Duration associated to the fuzzy value *HighDuration* (presented in fig. 1), the database engine
149 will first retrieve information corresponding to Duration value in the interval [50, 70] and

150 second, with a decreasing degree of preference, information corresponding to duration in the
151 interval [20, 50[or]70, 100].

152 Formally, for each element in a fuzzy set, there is a grade of membership score that represents
153 the degree to which the element belongs to the set. These scores are assumed to vary between
154 0 and 1 inclusive. A 0 indicates that the element has no membership in the set and a 1
155 indicates that the element has complete, or crisp, membership in the set.

156 It is a common think to find imprecise data in the bibliography, expressed as intervals of
157 value, without any reference to the manner it has been obtained. The reason why we have
158 chosen the possibility theory is that we want to represent imprecision that is characterized by
159 an inherent vagueness (as *high duration*) and not uncertainty which occurs when one can
160 accurately predict the behavior of many similar experiments (Laviolette et al., 1995). The
161 framework of fuzzy logic restated in the possibility theory base approach presented in Zadeh
162 (1978) has been used to represent precise or imprecise values of data and to realize fuzzy
163 pattern matching. A lot of works have focussed on the management of this kind of data in
164 databases (Dubois and Prade, 1995; Bosc and Pivert, 1997; Bosc and Pivert, 1998; Galindo et
165 al., 1998).

166 Representation of precise or imprecise values on continuous or discrete domain are illustrated
167 here with 3 examples. In Figure 1, the fuzzy set *HighDuration* represents an imprecise value :
168 the duration is completely possible in the interval 50-70 seconds, but it is still possible with a
169 decreasing degree of possibility until the interval 20-100 seconds. Figure 2 illustrates the
170 imprecise value *SterilizedMilk* which can be whole sterilized milk with a degree of possibility
171 of 1 and which may be also skimmed sterilized milk with a degree of possibility of 0.5. The
172 figure 3 shows that the same framework may also represent a precise value (here the precise
173 value 80 seconds for the attribute Duration), it may be used to represent both kinds of
174 information (precise or imprecise).

175 More formally, the value associated to an attribute A for a tuple t is defined as a possibility
 176 distribution $\pi_{A(t)}$ on the underlying domain D . The possibility distribution $\pi_{A(t)}$ can be viewed
 177 as a fuzzy restriction of the possible value of $A(t)$ by a fuzzy subset F . Given χ the
 178 membership function associated to F :

$$\begin{cases} \forall d \in D, \chi_F(d) \in [0,1] \\ \pi_{A(t)}(d) = \chi_F(d). \end{cases} \quad (1)$$

179 For instance, the information « *thermisation t has a high duration* » will be represented by:

180 $\pi_{Duration(t)}(d) = \chi_{HighDuration}(d)$. Here, $\chi_{HighDuration}$ is a membership function which represents
 181 the vague predicate *HighDuration* defined as a fuzzy subset. We consider two kinds of
 182 membership functions depending on the domain : (i) on an ordered and continuous underlying
 183 domain, a trapezoidal fuzzy subset is associated to each value (see fig. 1), (ii) on a discrete
 184 and non ordered domain, a standard fuzzy subset is associated to each value (see fig. 2).

185 **Statistical models**

186 Two types of models are used in order to give an estimation of the data sought. The first one
 187 concerns growth model and the second one inactivation model.

188 The growth model comes from a bacterial growth simulation model with two levels (Rosso et
 189 al., 1993). Primary level model F (adaptive model) describes how microbial density x changes
 190 with time t in a specified environment $x(t)=F(p_i, t)$ where p_i are at most 6 characteristic
 191 parameters (maximum growth rate μ_{max} or generation time $t_g=ln2/\mu_{max}$, lag phase duration λ ,
 192 inoculum x_0 , maximal density x_{max} , and eventually initial growth rate μ_0 , adaptation rate ρ_a).

193 Secondary level models indicate how parameters μ_{max} and λ of primary model change with
 194 respect to cardinal values (maximum, minimum and optimal) of temperature T and of pH and
 195 a fourth parameter μ_{opt} (value of μ at optimal temperature and optimal pH). With this model we

196 are able to give a fuzzy estimation of the parameter μ_{max} (or the generation time t_g) using fuzzy
197 arithmetic (Pedrycz and Gomide, 1998, pp129-150).

198 Inactivation models (Augustin, 1996) used are linear models : primary modeling is a linear
199 relation between \log_{10} (cell population N) and time t : $\log_{10}(N)=at+b$, allowing us to define
200 $D=1/a$, time of decimal reduction. Then a linear regression of $\log_{10}(D)$ on the temperature T :
201 $\log_{10}(D)=cT+d$, is calculated from numerous data found in the literature. This model allows us
202 to calculate the estimation of the duration D for a given temperature value and moreover to
203 calculate a fuzzy estimation of it. This calculus takes into account two kinds of variability :
204 the limits of the fuzzy subset representing the user's preferences and the confidence interval
205 on the coefficients of the regression model (Buche and Dervin, 1999).

206 Two conditions must be fulfilled to obtain a fuzzy estimation using those both models : (i)
207 values for each model parameter must be expressed in the selection attributes (for example,
208 temperature and pH for the selected growth model and temperature for the selected
209 inactivation model) ; (ii) values associated to each attribute expressed as fuzzy subset must be
210 included in the limits of the model. If the second condition is not fulfilled, we deliver a precise
211 estimation using precise values for model parameters (defined as the center of the core of the
212 trapezoidal fuzzy subsets).

213 However, our method is not dependent on those previous models because : first, our system
214 incorporates in its knowledge base predictive models which can be replaced or enhanced if
215 necessary; second, different models may be used for different categories of data (see section
216 The Information Retrieval System for more details).

217 **Resemblance relations**

¹ Defined as the set of values where the membership function is equal to 1

218 If no information stored in the database corresponds to the values specified for the selection
 219 attributes, it is still possible to retrieve the closest information if resemblance relations have
 220 been defined. Resemblance relations res_i are defined on the domain D_i of an attribute i which
 221 belongs to the query. Given a distance d_i on the domain D_i , the resemblance relations between
 222 x and y are defined as follows:

$$\begin{cases} res_i: D_i \times D_i \rightarrow [0,1] \\ \forall (x, y) \in D_i \times D_i, res_i(x, y) = \frac{1}{1 + d_i(x, y)}. \end{cases} \quad (2)$$

223 We have defined two kinds of distance d_i :

224 (i) for biological qualitative information (e.g., substrates or serotypes), called concepts in
 225 the following, we have defined a semantic distance d_i computed in
 226 generalization/specialization trees, called *GS* trees. A *GS* tree is defined as follows : concept
 227 C_1 is the father of C_2 (i.e. C_1 is above C_2 in the tree) if C_1 is more general than C_2 (C_2 is more
 228 specific than C_1). An example of *GS* tree is given in Fig. 4, representing the distance
 229 associated to the attribute Substrate. We drew upon the work of (Foo et al., 1989) to build a
 230 semantic distance function on *GS* trees. Given two concepts C_1 and C_2 , we consider C_3 , the
 231 more specific concept which generalizes C_1 and C_2 ; then the semantic distance $d_i(C_1, C_2)$ is
 232 equal to $d_i(C_1, C_3) + d_i(C_2, C_3)$. By example, in the fig.4, the distance between concepts
 233 *Skimmed milk* and *Whole milk* is 2 ($d_i(\text{Skimmed milk}, \text{Sterilized Milk}) + d_i(\text{Whole Milk},$
 234 $\text{Sterilized Milk}) = 1 + 1 = 2$) and the distance between *Skimmed milk* and *Raw milk* is 4 (d_i
 235 $(\text{Skimmed milk}, \text{Milk}) + d_i(\text{Raw Milk}, \text{Milk}) = 3 + 1 = 4$).

236 (ii) for information defined on an ordered domain, we have defined a distance d_i between
 237 two values x and y of an attribute i , as follows : given an increasing ordered series of values
 238 Δ_n ($\Delta_n < \Delta_{n+1}$) on the domain D_i associated to increasing numerical values α_n , $d_i(x, y) = \alpha_n$
 239 where α_n is the value associated to Δ_n such as $\Delta_n \leq |x-y| < \Delta_{n+1}$

240 By example, the following values α_n may be associated to the values Δ_n on the domain D_i of
 241 the attribute Temperature

n	1	2	3	4	...
Δ_n	0	1	3	5	...
α_n	0	1	2	3	...

242 Then for values 2,3,7 of attribute Temperature

243 $d_{\text{température}}(2, 3) = 1$ because $\Delta_2 \leq |2-3| < \Delta_3$

244 $d_{\text{température}}(3, 7) = 2$ because $\Delta_3 \leq |3-7| < \Delta_4$

245 **The information retrieval system**

246 Using a graphical user interface called GUI in the following (see fig 7), the user queries the
 247 database specifying selection criteria using the form attribute/value. The value associated to
 248 each selection attribute is a fuzzy subset. It represents the user's preferences. Result attributes
 249 are the closest values retrieved from the database for the selection attributes and specific
 250 attributes depending on the view through which the query is executed.

251 The system scans the database through the different views available in its knowledge base and
 252 retrieves the results associating an adequation degree to each of them. More precisely, the GUI
 253 translates and submits to the engine the user's preferences as a set of queries Q_V . For each
 254 available view V , the syntax of Q_V is a set $\{V, a_1, \dots, a_n, \langle a_{n+1}, v_{n+1}, r_{n+1}, ft_{n+1} \rangle, \dots, \langle a_m, v_m,$
 255 $r_m, ft_m \rangle, nb, t\}$ where V is the name of the view through which the query is executed; a_1, \dots, a_n
 256 are the expected attributes (limited to 1 in the examples), $\langle a_{n+1}, v_{n+1}, r_{n+1}, ft_{n+1} \rangle, \dots, \langle a_m, v_m,$
 257 $r_m, ft_m \rangle$ are quadruplets defining the selection criteria, nb and t define respectively the
 258 maximum number of tuples and the minimum degree of matching of each tuple in the result (t
 259 in $[0, 1]$). The quadruplets defining the selection criteria have the following meaning:

260 For all i in $[n+1, m]$:

- 261 • a_i is a selection attribute and v_i is a fuzzy set defining the value associated with a_i . We
- 262 distinguish two kinds of labels depending on the underlying domain of the attribute :
263 discrete or continuous. Two examples of fuzzy sets are given in fig 1 and fig 2.
- 264 • r_i is a boolean value specifying to the engine whether the resemblance relations
- 265 available in the engine knowledge base may be used or not.
- 266 • ft_i is a boolean value specifying to the engine whether the users' preferences expressed
- 267 in the fuzzy sets may be enlarged or not.

268 For instance, we consider the following query Q_{ex} :

269 $V=HeatResistance$

270 $a_1 = DecimalReductionTime,$

271 $\langle a_2=Substrate, v_2=MyEggProductPreferences, r_2= true, ft_2= false \rangle,$

272 $\langle a_3= DecimalReductionTime, v_3=HighDuration, r_3= false, ft_3= true \rangle$

273 $nb = 40, t = 0.3$

274 The value represented by the fuzzy subset $MyEggProductPreferences = \{(egg\ yolk, 1),$
275 $(whole\ egg, 0.5)\}$ indicates that the user is first interested by information about egg yolk and
276 second by whole egg. The definition of the fuzzy set is made by the user using the GUI. This
277 is the first way to enlarge the query , completely controlled by the user. The second way of
278 enlargement is also controlled by the user : he may activate or deactivate the automatic
279 enlargement based on the set of resemblance relations defined in the previous section and
280 stored in the knowledge base of the engine. The boolean value r_i , specified in the above
281 quadruplet definition, permits to control this mechanism. In the above example Q_{ex} , ($r_2= true$)
282 the engine is authorized to realize enlargement using the resemblance relation defined on the
283 domain of values of the attribute $Substrate$. The third way of enlargement is also controlled
284 by the user : he may activate or deactivate the automatic enlargement of his preferences based
285 on fuzzy tolerance coefficients defined in the knowledge base of the engine for each category

286 of data. The boolean value ft_i , specified in the above quadruplet definition, permits to control
287 this mechanism. In the above example Q_{ex} , ($ft_3 = true$) the engine is authorized to realize
288 enlargement on the user's preferences represented by the fuzzy set *HighDuration* for the
289 attribute *DecimalReductionTime*.

290 In the running version of the system prototype, the selection attributes available for each
291 view are substrate, conditions of the process step (temperature and pH) and bacterium. For the
292 moment, views available concern growth, heat resistance (inactivation) and bacteriocin
293 interaction. In this information retrieval system, we have introduced the new concept of
294 category-based fuzzy view (Buche and Loiseau, 1999; Buche and Dervin, 1999). This notion
295 is an extension of the classical view concept in databases, e.g. a virtual table in which all the
296 information needed by the user is brought together. This view is fuzzy because the selection
297 attributes are expressed as fuzzy predicates as in (Galindo et al., 1998; Zadrozny and
298 Kacprzyk, 1998). This view is category-based because the system identifies the category of
299 queried information (defined in the information retrieval system knowledge base). Three
300 advantages are provided by this category-based reasoning system implemented in the
301 category-based fuzzy view. First, if the user authorized it, the view engine enlarges user's
302 preferences expressed in his query if they are too restrictive due to the real content of the
303 database: it provides an additional degree of flexibility not available in the previous systems.
304 Second, the view engine is able to propose, in addition to the nearest information retrieved
305 from the database, an estimation of the data sought. For this purpose, statistical models well-
306 adapted to each category of data (choice of the model and associated parameters) are
307 incorporated into the knowledge base of the engine. Third, the view engine optimizes the
308 fuzzy matching processing because the user's preferences are only compared to the
309 information associated to the category selected : in the previous systems, all the information
310 stored in the queried tables are compared to the user's preferences.

311 Adequation between, on one hand, sought values and their category definition and, on the
312 other hand, sought values and values from the database are computed using pattern matching
313 techniques presented in (Dubois and Prade, 1995) and fuzzy object comparison measures
314 proposed in (Bosc and Pivert, 1997). A detailed description of the engine processing can be
315 find in (Buche and Loiseau, 1999).

316 We have implemented the concept of category-based fuzzy view in a prototype called CFQ
317 (for Category-based Fuzzy Querying). This software is written in Java language. Thus, CFQ is
318 available on the Web. The information is stored in an Oracle DBMS using the imprecision
319 representation model of FSQL (Galindo et al., 1998). A CFQ query is executed using a three
320 tier process architecture (see fig 8): (i) the CFQ Java client applet running under a usual Web
321 browser, (ii) a database connectivity middleware (running on the Web server due to security
322 checking on applet execution), (iii) an Oracle/FSQL database server. First, the user loads the
323 CFQ engine applet from the Web server on his Web browser. Second, he expresses his
324 preferences using the GUI of CFQ and asks for information scanning. Third, the GUI transfers
325 the user's query to the CFQ engine. The engine identifies the queried category of data and
326 builds the fuzzy filter. To retrieve the closest tuples stored in the database, the engine submits
327 SQL queries to the Oracle database server through the Web server.

328 **Results**

329 In this section, we present four results provided by our system CFQ corresponding to four
330 different examples of queries. Each query illustrates functionalities provided by the system :
331 in example 1, automatic category-based enlargement of data sought by the user to take into
332 account variability of data values stored in the database ; in the second and third examples,
333 comparison between estimated values and closest experimental values retrieved from the
334 database may provide on one hand a validation of the estimation or on the other hand

335 important differences which need to be explained by a sound bibliographical analysis ; in the
 336 fourth example, proposal of values unknown by the user but essential to obtain an estimation.
 337 The first example is about the behavior of *L. monocytogenes* in egg products under
 338 inactivation conditions. It is a sort of process validation query. The user specifies as querying
 339 attributes, substrate, temperature conditions and decimal reduction time of his process and
 340 wants to compare it to the decimal reduction time values stored in the database. Values sought
 341 for decimal reduction time are represented by the fuzzy set *HighDuration* (see fig 1). Values
 342 sought for temperature are in the interval [60, 64]. Values sought for substrates are
 343 represented by the following fuzzy set called *MyEggProductPreferences* = {(egg yolk, 1),
 344 (whole egg, 0.5)}. It means that the user is first interested by information about egg yolk and
 345 second by information about whole egg. The 3 views available in the system are executed and
 346 only the view *HeatResistance* provides results (with a adequation degree superior to a
 347 threshold, set arbitrarily to 0.3). The syntax of the query *Q* generated by the GUI
 348 corresponding to the view *HeatResistance* is the following :

349 $Q = \{V=HeatResistance, a_1 = DecimalReductionTime,$
 350 $\langle a_2=Substrate, v_2=MyEggProductPreferences, r_2= true, ft_2= false \rangle,$
 351 $\langle a_3= DecimalReductionTime, v_3=HighDuration, r_3= true, ft_3= true \rangle$
 352 $, nb = 40, t = 0.3\}$

353 In this view, the system has determined that the category of queried data is *Egg Product*. For
 354 this category, the system has found in its knowledge base an enlargement coefficient (called
 355 fuzzy tolerance coefficient) for the selection attribute *DecimalReductionTime*. Fig. 5 shows
 356 user's preferences enlargement for this selection attribute. The values sought for decimal
 357 reduction time are compared to the values stored in the database which are represented as
 358 imprecise values defined by the interval given by the two attributes *Duration Min* and
 359 *Duration Max*. Table 2 shows results obtained without or with enlargement of preferences

360 computed for the decimal reduction time attribute. The first five tuples appear in the two
 361 query results. An additional tuple is provided in the second result set thanks to the
 362 enlargement mechanism even if the threshold set on the adequation degree is the same for the
 363 two queries (set arbitrarily to 0.3 as said previously). This additional datum concerns a
 364 substrate which does not belong to the user's preferences expressed by
 365 *MyEggProductPreferences*. Therefore, it has been retrieved thanks to the fuzzy resemblance
 366 relation defined on the domain of values associated to the attribute Substrate and thanks to the
 367 category-based enlargement mechanism adapted to the category *EggProduct*. In the context of
 368 this example, it may be a relevant piece of information which indicates that the addition of salt
 369 or sugar to eggs provides a high increase in the decimal reduction time.

370 The second example is about the behavior of *L. monocytogenes* in dairy products under
 371 growth conditions. The user wants to retrieve the generation time and queries on substrate,
 372 temperature and pH attributes. Values sought are the following : for substrate, the fuzzy set
 373 *MyDairyProductPreferences* is {(Sterilized milk, 1)} which means that user queries sterilized
 374 milk only; for temperature, the fuzzy set *MyTemperaturePreferences* is the interval [2.2, 2.8]
 375 ; for the pH, the fuzzy set *MypHPreferences* is the precise value 6.6. The 3 views available in
 376 the system are executed and only the view *Growth* provides results. The syntax of the query
 377 *Q* generated by the GUI corresponding to the view *Growth* is the following :

378 $Q = \{V = \textit{Growth}, a_1 = \textit{GenerationTime},$
 379 $\langle a_2 = \textit{Substrate}, v_2 = \textit{MyDairyProductPreferences}, r_2 = \textit{true}, ft_2 = \textit{false} \rangle,$
 380 $\langle a_3 = \textit{Temperature}, v_3 = \textit{MyTemperaturePreferences}, r_3 = \textit{true}, ft_3 = \textit{false} \rangle,$
 381 $\langle a_4 = \textit{pH}, v_4 = \textit{MypHPreferences}, r_4 = \textit{true}, ft_4 = \textit{false} \rangle$
 382 $, nb = 40, t = 0.3\}$

383 In this case, the system CFQ is able to compute a fuzzified estimation thanks to the statistical
 384 growth model stored in the system knowledge base for the queried category *Dairy Product*.

385 The generation time estimation (in hours) computed with the sought values is the interval
386 [66.18, 105.27]. Simultaneously, CFQ system retrieves the closest experimental data stored in
387 the database (see table 3). The closest experimental value retrieved from the database
388 (identified by number 33) is out of the estimated interval of values (duration = 28.5 hours).
389 This value has been published in Walker et al. (1990), this bibliographical reference may also
390 be provided by the querying system. In this case, this experimental value tends to invalidate
391 the estimation.

392 The third example is about the behavior of *L. monocytogenes* in dairy products under
393 inactivation conditions. The user queries using substrate and temperature as selection
394 attributes. Values sought are the following : for substrate, the fuzzy set
395 *MyDairyProductPreferences* is {(Raw milk, 1)} ; for temperature, the value is defined by the
396 fuzzy set *ImpreciseTemperature* (see fig. 6). The 3 views available in the system are executed
397 and only the view *HeatResistance* (due essentially to the temperature values) provides results.
398 The syntax of the query *Q* generated by the GUI corresponding to the view *HeatResistance* is
399 the following :

400 $Q = \{V=HeatResistance, a_1 = DecimalReductionTime,$
401 $\langle a_2=Substrate, v_2= MyDairyProductPreferences, r_2= true, ft_2= false \rangle,$
402 $\langle a_3= Temperature, v_3= ImpreciseTemperature, r_3= true, ft_3= false \rangle$
403 $, nb = 40, t = 0.3\}$

404 As in the previous example, the CFQ system is also able to compute a fuzzified estimation
405 thanks to the statistical inactivation model stored in its knowledge base for the category of
406 queried data, here Dairy Products. The fuzzy kernel (values with a membership degree equal
407 to 1) of the estimated decimal reduction time (expressed in seconds) is the interval [4.83,
408 287.18]. The fuzzy support (values with a membership degree > 0) of the estimated decimal
409 reduction time is the interval [1.99, 697.99]. Most experimental values retrieved from the

410 database (see table 4) are included in the fuzzy estimation support. Only one experimental
411 value (ID = 8 in table 4) is far from the kernel (but still belongs to the support). Therefore, the
412 great majority of experimental values confirm the estimation.

413 The fourth example is about the behavior of *L. monocytogenes* in meat products under growth
414 conditions. In this case, we suppose that the user only knows the values for the substrate and
415 the temperature. He has no idea about the value of the pH. It is then possible to use the CFQ
416 system to perform an initial screening to retrieve possible values for the unknown parameter
417 (here the pH) and a second scan to obtain simultaneously an estimated value of the growth
418 (which can be computed only if temperature and pH are simultaneously provided in the query)
419 and a comparison with the closest experimental values.

420 For instance, in the first query, values sought are the following : for substrate, the fuzzy set
421 *MyMeatProductPreferences* is {(Cuts of beef, 1)} ; for temperature, the fuzzy set
422 *MyTemperaturePreferences* is the precise value 4°C. The 3 views available in the system are
423 executed and only the view *Growth* provides results (due to the temperature values). The
424 syntax of the query *Q* generated by the GUI corresponding to the view *Growth* is the
425 following :

426 $Q = \{V = \textit{Growth}, a_1 = \textit{pH}, a_2 = \textit{GenerationTime}$
427 $\langle a_3 = \textit{Substrate}, v_3 = \textit{MyMeatProductPreferences}, r_3 = \textit{true}, ft_3 = \textit{false} \rangle,$
428 $\langle a_4 = \textit{Temperature}, v_4 = \textit{MyTemperaturePreferences}, r_4 = \textit{true}, ft_4 = \textit{false} \rangle,$
429 $, nb = 40, t = 0.3\}$

430 The results of the interrogation are given in table 5. In the entire set of experiments, the pH is
431 included in the interval [5.5, 6.7]. In the second query, the user considers the same values for
432 substrate and temperature and adds the first screening for the pH value, i.e. the fuzzy set

433 *MyHPPreferences* corresponding to the interval [5.5, 6.7].). The syntax of the query *Q*
434 generated by the GUI corresponding to the view *Growth* is the following :

435 $Q = \{V = \textit{Growth}, a_1 = \textit{GenerationTime},$
436 $\langle a_2 = \textit{Substrate}, v_2 = \textit{MyMeatProductPreferences}, r_2 = \textit{true}, ft_2 = \textit{false} \rangle,$
437 $\langle a_3 = \textit{Temperature}, v_3 = \textit{MyTemperaturePreferences}, r_3 = \textit{true}, ft_3 = \textit{false} \rangle,$
438 $\langle a_4 = \textit{pH}, v_4 = \textit{MyHPPreferences}, r_4 = \textit{true}, ft_4 = \textit{false} \rangle$
439 $, nb = 40, t = 0.3\}$

440 The estimated generation time (expressed in hours) obtained during the second screening is
441 the interval [6.38, 76.3]. The closest set of experimental values is identical to the one obtained
442 in the first screening.

443 **Discussion**

444 One of the great difficulties in predicting microbial behavior in complex food matrix concerns
445 the fact that we have to take into account the variability of biological systems (bacteria and
446 food matrices) in a large range of environmental conditions (experimental or industrial
447 conditions). Moreover, in most cases, we are not able to determine all of the factors that could
448 interfere with the bacterial response. First, quantification cannot be done precisely (chemical
449 or physical gradients are not easy to establish), second, we do not always know the influence
450 of one factor (ingredient, other bacteria, physical conditions, etc.) on bacterial behavior. For
451 example, the ability of *L. monocytogenes* to grow in many food products is mainly evaluated
452 with regards to a small number of factors such as temperature, pH, etc. Nevertheless, other
453 factors (physiological, biological or other physical factors) could significantly modify the
454 growth response of *L. monocytogenes*. Because our knowledge in microbiology and
455 particularly in microbial ecology is growing, predictive microbiology models will be able to
456 take into account this new information in order to improve prediction. This new knowledge is
457 incomplete and/or can be imprecise. Our information retrieval system allows us to take into

458 account these data that can (i) modulate or specify the prediction obtained with mathematical
459 models or (ii) can propose a response when mathematical models have failed to calculate
460 growth or inactivation parameters.

461 We have also to mention two main drawbacks the user will have to manage with this new
462 querying approach. First, the adequation degree of a retrieved information may be the
463 combining result of three different enlargement mechanisms (user's preferences expression,
464 category-based enlargement of user's preferences, enlargement by resemblance relation
465 definition on the domain of values). Then, it is not easy to explain to the user how it has been
466 computed by the system. Second, the user has to be careful if he realizes comparisons between
467 closest experimental data retrieved from the database and fuzzy estimation using statistical
468 models. He has to check that retrieved experimental data are close enough compared to his
469 selection criteria, knowing that the adequation degree only provides a relative ranking of
470 retrieved information. For instance, in the result of the second query presented in the previous
471 section, only the first closest information (ID 33) checks completely the selection criteria and
472 can be really used to be compared to the fuzzified estimation.

473 **Conclusion**

474 In this paper, we have introduced the concept of category-based fuzzy view to design an
475 information retrieval system adapted to microbial risk analysis of food products. In the query,
476 the user has to specify his fuzzy preferences. Then the CFQ engine scans the database through
477 the views available in its knowledge base. For each view, the engine builds a fuzzy filter with
478 the user preferences *adapted to the category* of queried data - thanks to knowledge associated
479 to each category. It brings three benefits : (i) The CFQ engine *optimizes* the processing time
480 of the filtering because it only scans the data associated to the selected category ; (ii) the CFQ
481 engine can deliver an *estimate* of the queried data if an estimation method has been defined
482 for the selected category of data ; (iii) the CFQ engine can *enlarge* users' preferences to adapt

483 them to the real content of the database. The CFQ engine has been developed in Java
484 language. Therefore, the information retrieval system is available through the WEB. It has
485 been coupled to Oracle extended for fuzzy SQL (Buche and Loiseau, 1999; Buche and Dervin,
486 1999). CFQ is complementary with classical simulation software because :

- 487 - It can be used when classical software does not work; for example when parameters are
488 outside the limits of the model.
- 489 - It allows fuzzy expression of the parameters (selection attribute values) and takes into
490 account this imprecision in the database scan and in the calculation of the estimated
491 values.
- 492 - It allows comparison between estimated and experimental data, underlining possible
493 contradictions and then acting as a basic tool to understand unknown phenomena.
- 494 - It may provide information on an original domain, such as interaction between bacteria.
- 495 - It produces synthesized information, and provides bibliographical references to the
496 corresponding publications.
- 497 - It is a flexible tool able to evolve by incorporating new experimental results and/or new
498 simulation models.

499 In order to improve this information retrieval system, it is now necessary to add data that
500 come from more complete bibliography references. In the future, the database will have to
501 integrate other pathogens in order to carry out a better microbial prediction of food products,
502 predictive microbiology being an interesting tool for risk analysis..

503 **Acknowledgements**

504 This work was partially supported by the French Ministry of Agriculture's Food Department
505 (DGAL) and Research and Education Department (DGER).

506 **References**

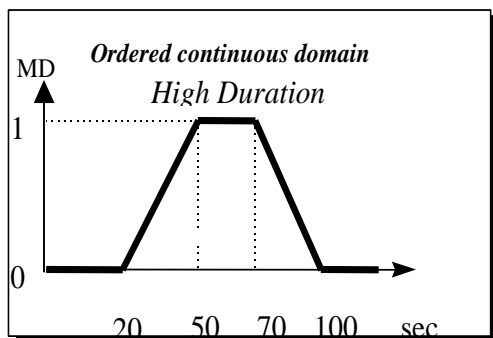
- 507 1. **Adair, C., P.A Briggs.** 1993. The concept and application of expert systems in the field of
508 microbiological safety. *J. of Industrial Microbiology.* 12:263-267.
- 509 2. **Alavi, S.H., V.M. Puri, S.J. Knabel, R.H. Mohtar, and R.C. Whiting.** 1999.
510 Development and validation of a dynamic growth model for *Listeria monocytogenes* in
511 fluid whole milk. *J.Food Prot.* 62:170-176.
- 512 3. **Augustin, J.C.** 1996. Résistance de *Listeria monocytogenes* aux traitements physiques.
513 *Path. Biol.* 44:790-807.
- 514 4. **Bajard, S., L. Rosso, G. Fardel, and J.P. Flandrois.** 1996. The particular behaviour of
515 *Listeria monocytogenes* under sub-optimal conditions. *Int. J. Food Microbiol.* 29:201-211.
- 516 5. **Baranyi, J., T.A. Roberts, and P. McClure.** 1993. A non autonomous differential
517 equation to model bacterial growth. *Food Microbiol.* 10:43-59.
- 518 6. **Baranyi, J., T.P. Robinson, A.Kaloti, and B.M. Mackey.** 1995. Predicting growth of
519 *Brochothrix thermosphacta* at changing temperature. *Int. J. Food Microbiol.* 27:61-75.
- 520 7. **Begot, C., I. Lebert, and A.Lebert.** 1997. Variability of the response of 66 *Listeria*
521 *monocytogenes* and *Listeria innocua* strains to different growth conditions. *Food*
522 *Microbiol.* 14:403-412.
- 523 8. **Bosc, P., and O. Pivert.** 1997. On the comparison of imprecise values in fuzzy databases.
524 *Proceedings of FUZZ-IEEE'97*, pp 707-712. Barcelona.
- 525 9. **Bosc, P., and O.Pivert.** 1998. A new approach to the filtering of little-known data.
526 *Proceedings of FUZZ-IEEE'98*, pp 1308-1313. Anchorage.
- 527 10. **Bosc, P., L. Liétard and O. Pivert.** 1999. Fuzzy theory techniques and applications in
528 database management systems. In *Fuzzy Theory Systems : Techniques and Applications*,
529 Vol.2, pp 631-686. Academic Press.

- 530 11. **Brouillaud-Delattre, A., M. Maire, C. Collette, C. Mattei, and C. Lahellec.** 1997.
531 Predictive microbiology of dairy products : influence of biological factors affecting growth
532 of *Listeria monocytogenes*. *J.AOAC Int.* 80:913-919.
- 533 12. **Buchanan, R.L., and J.G. Phillips.** 1990. Response surface model for predicting the
534 effects of temperature, pH, sodium chloride content, sodium nitrite concentration and
535 atmosphere on the growth of *Listeria monocytogenes*. *J. Food Prot.* 53:370-376.
- 536 13. **Buche, P., and S. Loiseau.** 1999. Using Contextual Fuzzy Views to Query Imprecise
537 Data. *Lecture Notes in Computer Science 1677. Proceedings DEXA'99 (Database and*
538 *EXpert system Application)* , pp 460-472. Florence, August 99.
- 539 14. **Buche, P. and C. Dervin.** 1999. Interrogation floue contextuelle de données imprécises,
540 application en microbiologie prévisionnelle. *Proceedings LFA'99 (Logique Floue et ses*
541 *Applications)*, pp 23-30. Valenciennes, Octobre 99.
- 542 15. **Buche, P., C. Dervin, A. Brouillaud-Delattre, and N. Gnanou-Besse.** 2000. A new tool
543 to combine experimental data and predictive microbiology in order to carry out the
544 microbial risk analysis of food products. In : Van Impe, J.F.M., Bernaerts, K. (Eds),
545 *Predictive Modelling in Foods- Conference Proceedings. KULeuven/Bio TeC, Belgium*
546 *(ISBN 90-804818-3-1)*, pp. 40-42.
- 547 16. **Codd, E.F.** 1970. A relational model for large shared data banks. *Communications of the*
548 *ACM.* Vol 13, N°6.
- 549 17. **Cox E. D,** 1995. *Fuzzy logic for business and industry.* Ed. Charles River Media. pp 113-
550 181.
- 551 18. **Dalgaard, P., and L.V. Jorgensen.** 1998. Predicted and observed growth of *Listeria*
552 *monocytogenes* in seafood challenge tests and in naturally contaminated cold-smoked
553 salmon. *Int. J. Food Microbiol.* 40:105-115.

- 554 19. **Dubois D., and H. Prade.** 1995. Tolerant Fuzzy pattern matching: an introduction, In:
555 Fuzziness in Database Management Systems (P. Bosc & J. Kacprzyk, eds.),
556 Heidelberg:Physica Verlag, 1995, pp. 42-58.
- 557 20. **Dubois, D., and H. Prade.** 1991. La logique floue. *La Recherche*.237: 1309-1315.
- 558 21. **Duh, Y., and W. Schaffner.** 1993. Modelling the effect of temperature on the growth rate
559 and lag time of *Listeria innocua* and *Listeria monocytogenes*. *J. Food Prot.* 56:205-210.
- 560 22. **Eifert, J.D., C. Gennings, W.H. Carter Jr., S.E. Duncan, and C.R. Hackney.** 1996.
561 Predictive model with improved statistical analysis of interactive factors affecting the
562 growth of *staphylococcus aureus* 196E. *J. Food Prot.* 59:608-614.
- 563 23. **Farrag, S.A., and E.B. Marth.** 1989. Growth of *Listeria monocytogenes* in the presence
564 of *Pseudomonas fluorescens* at 7°C or 13°C in skim milk. *J. Food Prot.* 52:852-855
- 565 24. **Fernandez, P.S., S.M. George, C.C. Sills, and M.W. Peck.** 1997. Predictive model of
566 the effect of CO₂, pH, temperature and NaCl on the growth of *Listeria monocytogenes*. *Int.*
567 *J. Food Microbiol.* 37:37-45.
- 568 25. **Foo, N., B.J. Garner, A. Rao, and E. Tsui.** 1989. Semantic distance in conceptual
569 graphs. Proceedings of the 4th annual workshop on conceptual structures. 1989. pp 1-9.
- 570 26. **Galindo J., J.C. Cubero, O. Pons, and J.M. Medina.** 1998. A server for fuzzy SQL
571 queries. Proceedings of the 1998 workshop FQAS'98 (Flexible Query-Answering
572 Systems), Roskilde, Denmark, pp 161-171, May 1998.
- 573 27. **Gay, M., O. Cerf, and K.R. Davey.** 1996. Significance of the preincubation temperature
574 and inoculum concentration on subsequent growth of *Listeria monocytogenes* at 14°C. *J.*
575 *Appl. Bacteriol.* 81:431-438.
- 576 28. **Guerzoni, M.E., R. Lanciotti, S. Torraiani, and F. Dellaglio.** 1994. Growth modelling
577 of *Listeria monocytogenes* and *Yersinia enterocolitica* in food model systems and dairy
578 products. *Int. J. Food Microbiol.* 24:83-92.

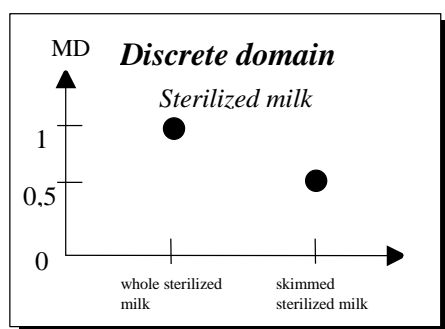
- 579 29. **Guinot-Thomas, P., M. Al-Ammoury, and F. Laurent.** 1995. Effect of storage
580 conditions on the composition of raw milk. *Int. Dairy J.* 5:211-223.
- 581 30. **Laviolette, M., J.W.J. Seaman, J.D. Baret, and W.H. Woodall.** 1995. A probabilistic
582 and statistical view of fuzzy methods. *Technometrics*.37-3 : 249-292.
- 583 31. **Mc Kellar, R.G., Buttler G., and K. Stanich.** 1997. Modelling the influence of
584 temperature on the recovery of *Listeria monocytogenes* from heat injury. *Food Microbiol.*
585 14:617-625.
- 586 32. **Mille, A., and A. Napoli.** 1997. Aspects du raisonnement à partir de cas. PRC-GDR
587 IA'97, pp 261-284.
- 588 33. **Murphy, P.M., M.C. Rea, and D. Harrington.** 1996. Development of a predictive model
589 for growth of *Listeria monocytogenes* in a skim milk medium and validation studies in a
590 range of dairy products. *J. Appl. Bacteriol.* 80:557-564
- 591 34. **Pedrycz, W., and F. Gomide.** 1998. An introduction to fuzzy sets, analysis and design.
592 MIT Press 1998.
- 593 35. **Ratkowsky, D.A., R.K. Lowry, T.A. McMeekin, A.N. Stokes, and R.E. Chandler.**
594 1983. Model for bacterial culture growth rate throughout the entire biokinetic temperature
595 range. *J. Bacteriol.* 154:1222-1226.
- 596 36. **Rosso, L., J.R. Lobry, and J.P. Flandrois.** 1993. An unexpected correlation between
597 cardinal temperatures of microbial growth highlighted by a new model. *J. Theor. Biol.*
598 162:447-463.
- 599 37. **Rosso, L.** 1995. Modélisation et microbiologie prévisionnelle : élaboration d'un nouvel
600 outil pour l'agro-alimentaire. Thèse de l'Université Claude-Bernard – Lyon I. 1195.
- 601 38. **Salotti, S.** 1992. Filtrage flou et représentation centrée-objet pour raisonner par analogie:
602 le système FLORAN. Thèse de l'Université Paris XI Orsay. December 1992.

- 603 39. **Siswanto, H.P., and J. Richard.** 1992. Vitesse de croissance dans le lait de *Listeria*
604 *monocytogenes* et autres souches du même genre à des températures sub-optimales. Lait
605 72:265-275.
- 606 40. **Van Impe, J.F., B.M. Nicolai, M. Schellekens, T. Martens, and J.D. Baerdemaker.**
607 1995. Predictive microbiology in a dynamic environment : a system theory approach. Int.
608 J. Food Microbiol. 25:227-249.
- 609 41. **Walker, S.J., P Archer, and J.G. Banks.** 1990. Growth of *Listeria monocytogenes* at
610 refrigeration temperatures. Journal of Applied Bacteriology. 68 :157-162.
- 611 42. **Zadeh, L.A.** 1965. Fuzzy sets. Information and control, n°8, pp 338-353.
- 612 43. **Zadeh, L.A.** 1978. Fuzzy sets as basis for a theory of possibility. Fuzzy Sets and Systems.
613 1 :3-28.
- 614 44. **Zadrozny, S., and J. Kacprzyk.** 1998. Implementing fuzzy querying via the
615 internet/WWW: Java applets, activeX controls and cookies. Proceedings of the 1998
616 workshop FQAS'98 (Flexible Query-Answering Systems), Roskilde, Denmark, pp 358-
617 369, May 1998.
- 618 45. **Zwietering, M.H., J.T. de Koos, B.E. Hasenack, J.C. de Wit, and K. Van't Riet.** 1991.
619 Modeling of bacterial growth as a function of temperature. Appl. Environ. Microbiol.
620 57:1094-1101.
- 621 46. **Zwietering, M.H., T. Wiltjes, J.C. de Wit, and K. Van't Riet.** 1992. A Decision
622 support system for Prediction of the Microbial Spoilage in Foods. J. Food Prot. 55:973-
623 979.
- 624



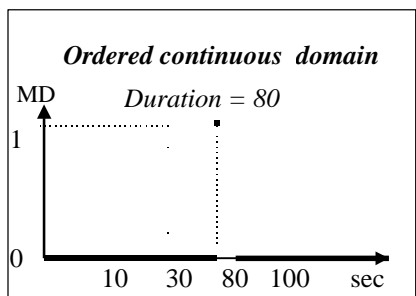
625
626
627
628

Figure 1. A trapezoidal fuzzy subset representing an imprecise value
(MD stands for Membership function)



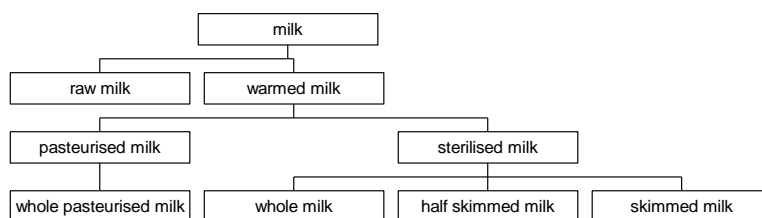
629
630
631

Figure 2. A fuzzy subset defined on a discrete domain



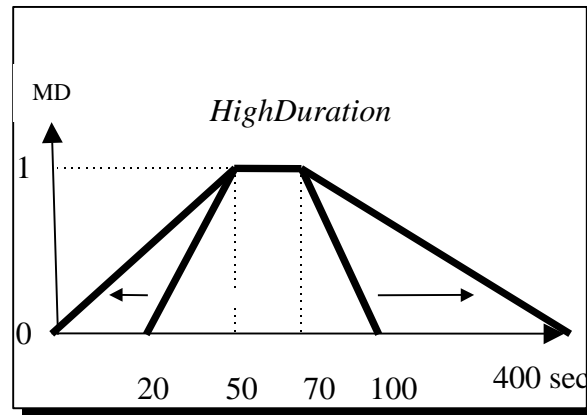
632
633
634

Figure 3. A fuzzy subset representing a crisp (precise) value



635
636
637
638

Figure 4. Example of Generalization-Specialization (GS) tree for milk products

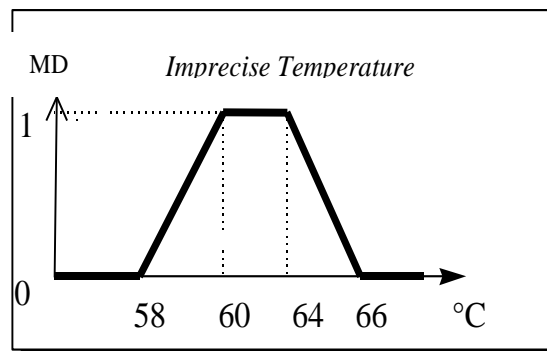


640 **Figure 5. Enlargement mechanism of the user's preferences using fuzzy tolerance coefficient**

641 **for the category EggProduct (first example of query)**

642

643

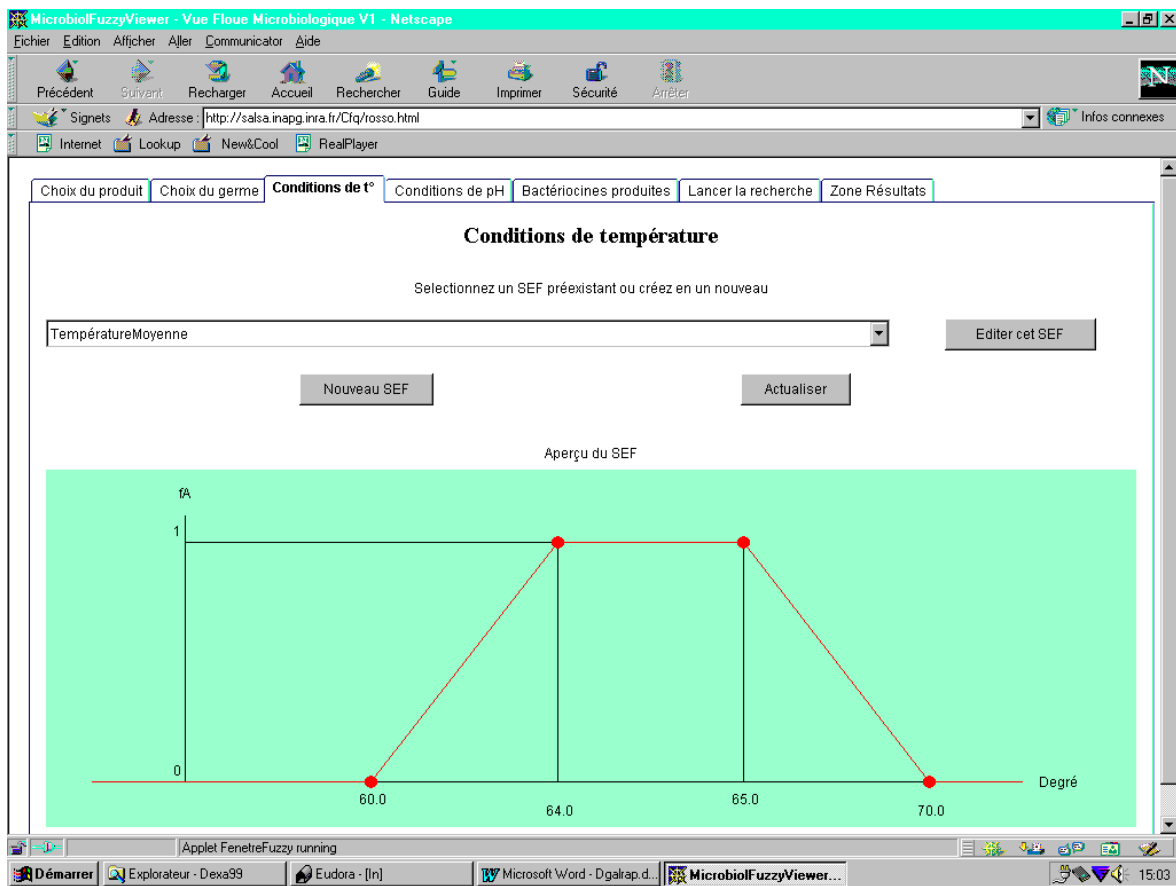


644

645 **Figure 6. Temperature value for the third query**

646

647



649

650 **Figure 7. The Graphical User Interface permits to express the selection criteria of a query and to**
 651 **display the results. It is organized as a thumbnail formular. In the above example, the selected thumbnail**
 652 **shows the value (expressed as a fuzzy set) associated by the user to the attribute temperature.**

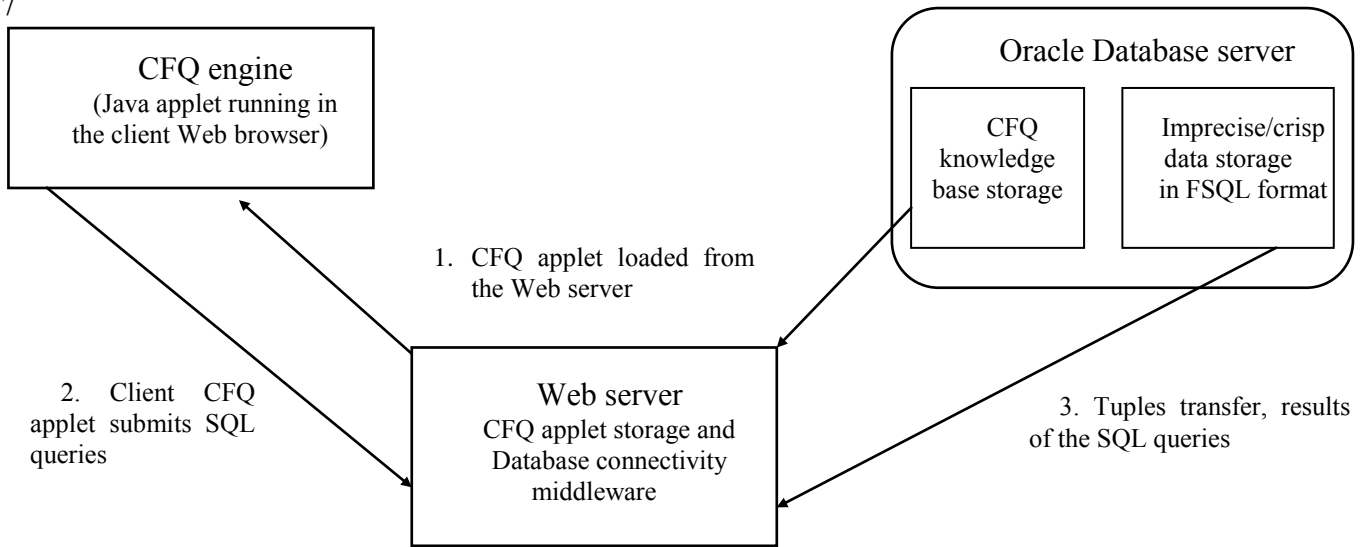
653

654

655

656

657



658

659 **Figure 8. Category-based Fuzzy Querying engine is a software based on a three tier architecture : the**
660 **Java applet is available on the Web server, the Java applet is loaded on the client to run in the Web**
661 **browser, the CFQ engine running in the client applet queries the database through the Web server to**
662 **retrieve the closest information**

663

664

665

666

Table 1. View BacteriocinProducedByBacterium

BACTERIOCIN NAME	BACTERIUM NAME	STRAIN NAME
Pediocin ACH	Lactobacillus Plantarum	WHE92
Pediocin ACH	Pediococcus Acidilacti	H

667

668

669

670 **Table 2. Closest data retrieved through the view *Heat Resistance* corresponding to the first query without**
671 **enlargement of preferences (Adequation Degree (AD) > 0.3) are the first five tuples. With enlargement of**
672 **preferences, same first five tuples are obtained plus the additional sixth tuple.**
673

ID	AD	SUBSTRATE	TEMP (°C)	DURATION MIN (Sec)	DURATION MAX (Sec)
80	0.905	egg yolk	60	55.2	101.4
82	0.905	egg yolk	60	55.2	101.4
85	0.786	egg yolk	64.4	26.4	26.4
78	0.5	whole egg	60	78	84
81	0.5	whole egg	60	90	126
86	0.319	egg + salt or sugar	64.4	366.6	366.6

674

675

676

677

**Table 3. Closest data retrieved through the view *Growth* corresponding to the second query
(AD means Adequation Degree)**

678

679

ID	AD	SUBSTRAT	TEMP	PH	GENERATION TIME (hours)
33	0.76	Sterilized_milk_(UHT)	2.5	6.6	28.5
4	0.35	Skimmed_milk	4	6.4	32.5
7	0.35	Skimmed_milk	4	6.4	30.6
10	0.35	Skimmed_milk	4	6.4	37.5
13	0.35	Skimmed_milk	4	6.4	38
16	0.35	Skimmed_milk	4	6.6	48.9
17	0.35	Skimmed_milk	4	6.6	53.5
20	0.35	Sterilized_whole_milk	4	6.4	30.6
23	0.35	Sterilized_whole_milk	4	6.4	30.5
26	0.35	Sterilized_whole_milk	4	6.4	36
29	0.35	Sterilized_whole_milk	4	6.4	36.5

680

681

682

683

Table 4. Closest experimental values for the third query

684

(AD means Adequation Degree)

685

686

ID	AD	SUBSTRATE	TEMP	DURATION MIN (sec)	DURATION MAX (sec)
9	0.76	Raw_milk	63.3	19.8	19.8
8	0.5	Milk	61.7	570	570
11	0.2	Raw_milk	71.7	1.6	1.6
12	0.2	Raw_milk	71.7	1.9	1.9
13	0.2	Skimmed_milk	62.7	54	54
14	0.2	Sterilized_whole_milk	62.7	60	60
15	0.2	Half_skimmed_milk	62.7	54	54
16	0.2	Half_skimmed_milk	62	6	24
17	0.2	Half_skimmed_milk	62	6	24
18	0.2	Sterilized_whole_milk	63.3	19.8	34.8
19	0.2	Skimmed_milk	63.3	25.8	28.2
21	0.2	Sterilized_whole_milk	71.7	3.1	3.1
22	0.2	Sterilized_whole_milk	71.7	5	5
23	0.2	Half_skimmed_milk	72	0.05	1.5
24	0.2	Half_skimmed_milk	72	0.02	1.4
25	0.2	Sterilized_whole_milk	71.7	2.7	2.7
26	0.2	Sterilized_whole_milk	71.7	1.3	1.3
27	0.2	Sterilized_whole_milk	71.7	0.7	0.7
28	0.2	Sterilized_whole_milk	71.7	2	2
29	0.2	Sterilized_whole_milk	71.7	1.4	1.4
30	0.2	Sterilized_whole_milk	71.7	0.6	0.6
31	0.2	Sterilized_whole_milk	71.7	1.2	1.2
32	0.2	Pasteurized_whole_milk	57.8	138	138

687

688

689

690

Table 5. Closest experimental values for the fourth query

691

(AD means Adequation Degree)

692

693

ID	AD	SUBSTRATE	TEMP	PH	GENERATION TIME (hours)
86	0.58	Cuts_of_fat_beef	4.7	5.6	11.9
81	0.5	Cuts_of_lean_beef	5	6.3	17.1
79	0.4	Cuts_of_lean_beef	5.5	5.5	22.1
77	0.38	Cuts_of_lean_beef	5.6	5.6	22.9
87	0.24	Cuts_of_fat_beef	7.5	5.6	7.1
85	0.21	Cuts_of_lean_beef	10	6.7	4.08
88	0.20	Cuts_of_fat_beef	10.1	5.6	4.62

694