HAL open science

# Towards the Automatic Detection of Nutritional Incompatibilities Based on Recipe Titles

Nadia Bebeshina-Clairet, Mathieu Lafourcade

## To cite this version:

Nadia Bebeshina-Clairet, Mathieu Lafourcade. Towards the Automatic Detection of Nutritional Incompatibilities Based on Recipe Titles. CD-MAKE: Cross-Domain Conference for Machine Learning and Knowledge Extraction, Aug 2017, Reggio, Italy. pp.346-366, 10.1007/978-3-319-66808-6_23 . hal-01677134

**HAL Id: hal-01677134**

**https://inria.hal.science/hal-01677134v1**

Submitted on 8 Jan 2018

**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.

Distributed under a Creative Commons Attribution 4.0 International License

# Towards the automatic detection of nutritional incompatibilities based on recipe titles

Nadia Clairet
LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier) / 860 rue de St Priest 34000 Montpellier, France
LIMICS (Laboratoire d'Informatique Mdicale et d'Ingnieurie des Connaissances en e-Sant) / 74 rue Marcel Cachin 93017 Bobigny, France
Lingua et Machina / 7 Boulevard Anatole France, 92100 Boulogne-Billancourt, France
clairet@lirmm.fr
Mathieu Lafourcade
LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier) / 860 rue de St Priest 34000 Montpellier, France
mathieu.lafourcade@lirmm.fr

**Abstract.** The present paper reports experimental work on the automatic detection of nutritional incompatibilities of cooking recipes based on their titles. Such incompatibilities viewed as medical or cultural issues became a major concern in western societies. The gastronomy language represents an important challenge because of its elusiveness, its metaphors, and sometimes its catchy style. The recipe title processing brings together the analysis of short and domain-specific texts. We tackle these issues by building our algorithm on the basis of a common knowledge lexical semantic network. The experiment is reproducible. It uses freely available resources.

## 1 Introduction

The analysis of cooking recipes is a very challenging task when it comes to automatically detect the compatibility of a dish with a diet. Indeed, performing such detection given the dish title as it can be found in a restaurant menu implies solving the following issues:

- short text analysis : how to overcome the context scarcity?
- domain specific text analysis : how to select the relevant information for the processing?
- qualified output structure : how to reflect the shades of the nutritional incompatibility as it may be strict or calibrated (forbidden, undesirable, small quantity authorized, fully authorized ingredients)?

We shall tackle these issues by immersing domain specific knowledge into a large general knowledge lexical semantic network, then by building our algorithm on top of it. In terms of structure, the network we use is a directed graph where nodes may represent simple or compound terms, linguistic information, phrasal expressions, and

sense refinements[1]. The arcs of the graph are directed, weighted, and typed according to the ontological, semantic, lexical associations between the nodes. They also may be semantically annotated (e.g. *malignant tumor* $\xrightarrow{characteristic/frequent}$ *poor prognosis*) which is useful for working with domain specific expert knowledge. During the traversal (interpretation) of the graph, the nodes and the arcs are referred to as respectively *terms* and *relationships*. Thus, in our discourse, a relationship is a quadruplet $R = \{term_{source}, type, weight, term_{target}\}$. The weight can be understood as an association force of the relationship between two terms of the network.
The aim of the present graph browsing experience is to obtain probabilistic incompatibility scores given a list of raw recipe titles and a lexical semantic network (directed, weighted, and typed graph). The paper will be structured as follows. First, we will evoke main state-of-the-art achievements related to short text analysis, cooking recipe analysis, and specific applications for the nutrition domain. Second, we will introduce our experimental setup. Third, we will describe our method. Finally, we will present and discuss the output of our system and sketch some possible evolutions and applications of the method.

## 2 State of the Art

Recipe titles can be viewed as short texts on a particular domain. In terms of semantic information required for the analysis and methodology, the processing of recipe titles can be considered as a possible specialization of some general method for short text analysis. Two approaches to short text processing will be highlighted in this section: knowledge based distributional method and logic based distributional method. The methods proposed for the analysis of domain specific texts (cooking recipes) are also relevant as they point out the selection of relevant information for domain specific text analysis. We will detail the main approaches to the analysis of cooking recipes and conclude this section by discussing the domain specific systems and the way the cooking recipes can be represented in order to facilitate their semantic processing.
An interesting example of a knowledge-based method for the short text analysis is the method developed by[11] in the framework of the Probase[2] lexical semantic network. As a knowledge resource, Probase is positioned as a graph database helping to better understand human communication and resolve ambiguities related to the human common sense knowledge. The Probase-based tools allow obtaining a concept distribution based on different scoring functions (conditional probability distribution[3], point-wise

[1] Unlike the dictionary sense, the sense refinement reflects the typical use of a term, its sense activated by a particular context
[2] `https://www.microsoft.com/en-us/research/project/probase/`
[3] Conditional probability distribution of two discrete variables (words)

$$P(A|B) = \frac{P(A) \cap (B)}{P(B)}$$

mutual information[4] etc.). In terms of relationship type representation, the publicly available Probase release only provides the *is-a* relationship data mined from billions of web pages. In the recent years, the research experiments using Probase has been focusing on segmentation, concept mapping, and sense disambiguation. As part of this work, [11] introduced a method for short text analysis that has been tested for analyzing queries. Prior to the analysis of short texts, the authors acquire knowledge from web corpus, Probase network as well as a verb and adjective dictionary It is stored as a set of resources such as co-occurrence network, *is-a* network, concept clusters, vocabularies etc. They solve the segmentation task by introducing a multi-fold heuristic for simple and multi-word term detection. It takes into account the presence of *is-a* and *co-occurrence* relationships between the candidate terms. Then, the terms are typed according to different categories : part of speech, concept/entity distinction etc. Finally, the disambiguation is done using the weighted vote (conceptual connections of the candidate term considering the "vote of context"). This method seems to be relevant for queries, however it would be difficult to apply it for the analysis of recipe titles and the detection of nutritional incompatibilities. The main difficulty comes from the fact that it is concept driven. Indeed, for a term such as "creme brulee", we obtain the concept distribution scores shown in the Table 1. Due to underlying semantic relationship

| Term (concept) | Score by MI | Score by $P(concept\|entity)$ |
|---|---|---|
| dessert | 0.386 | 0.446 |
| authentic bistro dish | 0.164 | 0.12 |
| off beat flavor | 0.047 | 0.06 |
| homemade dessert | 0.046 | 0.048 |
| dairy based dessert | 0.04 | 0.036 |

**Table 1.** Concept distribution for *creme brulee* (Probase conceptualization tool)

types (*is-a*; *gloss* or *relatedness*), these examples bring very few information about the composition and the semantic environment (that needs considering relationship types expressing *part-whole*, *location*, *instrument*, *sensory characteristic* relationships) of the recipes and one can hardly qualify some of the returned scores (e.g. *off beat flavor*) in order to approximate the underlying relationship type. An improvement of the knowledge based distributional method could be made using an additional semantic resource containing a rich part-whole semantics[5] such as WordNet [9]. This type of resource could be used as a reference for a part-whole relationship discovery from large web corpus. There has been a number of proposals in the recent years and among them the semi-

[4] Point-wise Mutual Information is calculated according to the formula :

$$PMI(x;y) = \log \frac{P(x,y)}{P(x)P(y)}$$

[5] The WordNet part-whole relation splits into three more specific relationships : *member-of*, *stuff-of*, and *part-of*

automatic approach proposed by [10]. A different kind of methods for short text analysis may rely on general-purpose first order probabilistic logic as shows the approach developed by [4]. In the framework of this hybrid method, the distributional semantics is used to represent the meaning of words. Some of the lexical and semantic relations such as synonymy and hyponymy can be predicted. The authors use them to generate an on-the-fly ontology that contains only the relevant information related to some current semantic analysis task. They argue that the first-order logic has a binary nature and thus cannot be graded. Therefore, they adopt probabilistic logic as it allows weighted first order logic formulas. The weight of the formulas corresponds to a certainty measure estimated from the distributional semantics. First, natural language sentences are mapped to a logical form (using the Boxer tool [5][6]). Second, the ontology is built and encoded in the form of weighted inference rules describing the semantic relations (the authors mention *hyponymy*, *synonymy*, *antonymy*, *contextonymy* i.e. the relation between "hospital" and "doctor"). Third, a probabilistic logic program answering the target task is created. Such program contains the evidence set, the rule base (RB, weighted first order logical expressions), and a query. It calculates the conditional probability $P(Query|Evidence, RB)$. This approach has been tested on the SICK[7] data for the tasks of semantic textual similarity detection and textual entailment recognizing. The results showed a good performance of this method over distributional only and logic only methods. This kind of approach rather considers lexical relations (revealing some linguistic phenomena), than purely semantic (language independent, world knowledge phenomena) relations. As the knowledge based distributional method, it suffers from the difficulty to qualify the relationships as it uses mainly the co-occurrence analysis.
The analysis of cooking recipes is a flourishing research area. Besides the approaches focused on flavor networks which consider cooking recipes as "bags of ingredients" with a remarkable contribution of [1], the existing approaches to the recipe analysis concentrate on recipes taken as a sequence of instructions that can be mapped to a series of actions. Numerous publications report the implementation supervised learning methods. For instance, [18] use the annotated data to extract predicate-argument structures from cooking instructions in Japanese in order to represent the recipe as a work flow. The first steps of this process are words segmentation and entity type recognition. The latter is based on the following entity types: Food, Quantity, Tool, Duration, State, chefs action, and foods action. Therefore, this task is similar to the conceptualization process proposed by [11] in the framework of knowledge based short text analysis and discussed earlier in this section. Entity type recognition is followed by syntactic analysis that outputs a dependency tree. The final step aims at extracting predicate-argument triples from the disambiguated (through segmentation, entity type recognition, and dependency parsing) recipe text. In this approach, the semantic information that could be

[6] Boxer [5] is a semantic parser for English texts based on Discourse Representation Theory

[7] *Sentences Involving Compositional Knowledge*. This dataset includes a large number of sentence pairs that are rich in the lexical, syntactic and semantic phenomena (e.g., contextual synonymy and other lexical variation phenomena, active/passive and other syntactic alternations, impact of negation, determiners and other grammatical elements). Each sentence pair was annotated for relatedness and entailment by means of crowd-sourcing techniques. `http://clic.cimec.unitn.it/composes/sick.html`

attached to the arcs is attached to the nodes. The node type together with the syntactic markers (i.e. case marker) helps determining the nature of the predicate argument relation. This method yields modest results and could be improved by using more annotations and also by adopting a more versatile graph structure (i.e. a structure with typed arcs). [13] proposed a similar approach as part of an unsupervised technique for mapping recipe instructions to actions based on a hard Expectation Maximization algorithm and a restricted set of verb argument types (*location*, *object*).
In the paradigm of semantic role labeling, the approach of [17] use a Markov decision process where ingredients and utensils are propagated over the temporal order of instructions and where the context information is stored in a latent vector which disambiguates and augments the instruction statement under analysis. In this approach, the context information corresponds to the state of the kitchen and integrates the changes of this state according to the evolving recipe instructions. The changes are only partially observed (the authors assume that some instruction details may be omitted in the recipe text) and the resulting model is object-oriented. Each world state corresponds to a set of objects (i.e.ingredients, containers) along with predicates (quantity, location, and condition) of each object. Each action of the process is represented by a verb with its various arguments (semantic roles). Such representation indicates how to transform the state. The model also uses a simple cooking simulator able to produce a new state from a stream of low-level instructions to reflect the world dynamics model $p(State_t|State_{t-1}, Action_t)$.
In the case based reasoning paradigm, [8] represent the cooking instructions as a workflow and thus propose a method for the automatic acquisition of a rich case representation of cooking recipes for process-oriented case-based reasoning from free recipe text. The cooking process is represented using the Allen [3] algebra extended with relations over interval durations. After applying classical NLP tools for segmentation, part-of speech tagging, syntactic analysis, the extraction process from texts focuses on the anaphora resolution and verb argument analysis. The actions are modeled on the basis of the instructional text *without* considering the implicit information proper to the cooking recipes. The underlying knowledge resource is the case-based reasoning system Taaable [7].
Despite the variety of their theoretical background, the existing methods of cuisine texts analysis converge on the necessity to have a richer context around the terms present in the input text. Such context can be obtained using the following approaches:

– specific meta-language such as MILK[8], proposed as a part of the SOUR CREAM project [23], SIMMR [12] [9]. SOUR CREAM stands for "System to Organize and Understand Recipes, Capacitating Relatively Exciting Applications Meanwhile". MILK has been proposed as a machine-readable target language to create sets of instructions that represent the actions demanded by the recipe statements. It is is based on first-order logic, but allows handling the temporal order as well as creation/deletion of ingredients. A small corpus of 250 recipes has been manually annotated using MILK (CURD (Carnegie Mellon University Recipe Database)). Similarly, SIMMR [12] allows to represent a recipe as a dependency tree. The

[8] *Minimal Instruction Language for the Kitchen*
[9] *Simplified Ingredient Merging Map in Recipes*

leaves of the tree are the recipe ingredients and its internal nodes are the recipe instructions. Such representation supports semantic parsing. MILK tags have been used to construct the SIMMR trees. The authors also propose a parser to generate SIMMR trees from the raw recipe text. Machine learning methods (SVM classification) have then been used for instruction-ingredient linking, instruction-instruction linking using SIMMR;
- dynamic structures such as latent vector used by [17] and described earlier in this section;
- graph-shaped resources (ontologies, semantic networks) and their use by projection on the text under scope (as described earlier in this section) .

Among the projects centered on building specific resources and applications for the food and nutrition domain, the PIPS[10] project and the OASIS[11] project appeared as pioneering large-scale European ventures dedicated to the promotion of healthy food practices and to the building of counseling systems for the nutrition field. The PIPS project proposed a food ontology for the diabetes control whereas the OASIS project focused on nutritional practices of elderly people. Later on, some work has been centered on menu generation considered as a multi-level optimization issue (MenuGene [20]) [12]. Other authors introduced the proposal of alternative menus (Semanticook [2]). The case-based reasoning Taaable [7] [13] system evolved in the same direction and has been provided with nutritional values as well as nutritional restrictions and geographical features related to the recipes stored in the case base. However, in this paradigm, a formal representation of recipes in terms of semantic properties (directed acyclic graph) is mandatory. The Taaable knowledge repository and the formal framework of this system allow the representation of incompatibilities based on nutritional values and the subsumption relation i.e. *alcohol free*, *cholesterol free*, *gluten free*, *gout*, *vegan*, *vegetarian*, *nut free*.
We briefly introduced the main approaches capable of handling recipe title analysis. Nevertheless, to our knowledge, there has been no proposals of a system designed for recipes or recipe titles analysis in order to determine nutritional incompatibilities against a list of diets (medical or cultural). In the next section, we will propose this kind of system based on a general knowledge lexical semantic network.

## 3 Experimental setup

### 3.1 Problem statement

The problem can be stated as follows: how to detect nutritional restrictions from raw data using a background knowledge resource (lexical semantic network), rank the obtained incompatibility scores, and maximize the confidence about the scores. We leave the last point for future work.

[10] `http://cordis.europa.eu/project/rcn/71245\_en.html`. Concepts such as *diet*, *nutrient*, *calories* have been modeled in the framework of this project.
[11] http://www.oasis-project.eu/
[12] http://www.menugene.com/
[13] `http://wikitaaable.loria.fr`

### 3.2 Nutritional incompatibility representation

For our experiment, we use the recipe representation shared by the recipe content owners and available at schema.org[14]. In the scope of technical and semantic interoperability, we focus on the property *suitableForDiet* which concerns the concept *Recipe* and, subsequently the type *RestrictableDiet* and its instances defined as follows: *DiabeticDiet*, *GlutenFreeDiet*, *HalalDiet*, *HinduDiet*, *KosherDiet* , *LowCalorieDiet*, *LowFatDiet*, *LowLactoseDiet*, *LowSaltDiet*, *VeganDiet*, *VegetarianDiet*. Thus, our results may be easily encoded in the widely shared format in order to enhance the existing recipe representation. For these diets, we automatically extracted and manually validated an initial set of approximately 200 forbidden ingredients. We used domain-specific resources (lists of ingredients) found on the Web for this step. These ingredients (if not already present in our graph-based knowledge resource) have been encoded as nodes and linked to the nodes representing the diets listed above by the arcs typed *r_incompatible*. The nutritional restrictions differ in terms of their semantic structure (which may rely on nutritional composition (*part-whole* relation), cutting types (*holonymy* etc.), cooking state (i.e. boiled carrots are undesirable in case of diabetes). Basically, the incompatibility detection is a non linear classification problem as one recipe may be incompatible with several diets and with a different "degree" of incompatibility. Therefore, it needs a qualifying approach as in some cases (i.e. diabetes, low salt) a food may be not strictly forbidden but rather taken with caution.

### 3.3 Corpus

For our experiment we used a set of 5 000 recipe titles in French, which corresponds to a corpus[15]. of 19 000 words and a vocabulary of 2 900 terms (after removing stop words and irrelevant expressions such as "tarte tatin *à ma façon*", *tarte tatin my way*). This data mirrors the French urban culinary tradition as well as the established practice of searching the Web for recipes. The most important terms (and their number of occurrences in the corpus) are salade *salad* (157); tarte, poulet *tart, chicken* (124); soupe *soup* (112); facile *easy* (93); chocolat *chocolate* (89); saumon, gâteau *salmon, pastry* (87); légume, confiture *vegetable, jam* (86). These occurrences highlight the terms for which we need to have more details in our knowledge resource to perform the analysis.

### 3.4 Knowledge resource

**Why using a large general knowledge resource?** The data scarcity proper to the recipe titles is an important obstacle to their processing. If we focus on a plain co-occurrence analysis and compare our corpus to the list of forbidden ingredients, we only obtain 998 straightforward incompatibilities i.e. for which the forbidden ingredient is part of the recipe title and no specific analysis is needed to detect it. Distributional scores used by [11] and [4] may be interesting in terms of flavor associations, but they also demonstrate that we need to know more about the most likely semantic neighborhood of the term to handle the part-whole semantics and incompatibility analysis.

[14] schema.org/Recipe

[15] The corpus has been collected from the following Web resources : *15%* `www.cuisineaz.fr`*, 20%* `www.cuisinelibre.org` *et 65%* `www.allrecipe.fr`

**Enhancing the knowledge resource for the analysis** The semantic resource we use for the experiments is the RezoJDM lexical semantic network for French [16]. This resource stems from the game with a purpose AI project JeuxDeMots [14]. Built and constantly improved by crowd-sourcing (games with a purpose, direct contribution), RezoJDM is a directed, typed, and weighted graph. Today it contains 1.4M nodes and 90M relations divided into more than 100 types. The structural properties of the network have been detailed in [15] and later in [6], its inference and annotation mechanisms respectively by [24] and [22]. The ever ending process of graph population[17] is carried on using different techniques including games with a purpose, crowd-sourcing, mapping to other semantic and knowledge resources such as Wikipedia or BabelNet[19]. In addition, endogenous inference mechanisms, introduced by [24] are also used to populate the graph. They rely on the transitivity of the *is-a*, *hyponym*, and *synonym* relationships and are built for handling polysemy. In addition to the hierarchical relation types (*is-a* relation, *part-whole* relations), RezoJDM contains grammatical relations (*part-of-speech*), causal, thematic relations as well as relations of generativist flavor (ex. *telic role*). This resource is considered as a closed world i.e. every information that is not present in the graph is assumed as false. Therefore, the domain specific subgraph of the resource related to food and nutrition has been populated in order to allow cuisine text analysis.

It has been demonstrated by [22] that, for the sake of precision, general and domain specific knowledge should not be separated. Thus, for our experiment we do not build any specific knowledge resource dedicated to the recipe analysis. Instead, we immerse nutrition, sensory, technical knowledge into RezoJDM to enhance the coverage of the graph. This has been done partly through the direct contribution. External resources such as domain specific lexicons and terminological resources have been used. In particular, some equivalence and synonym relation triples have been extracted from IATE[18] term base[19]. The Agrovoc[20] thesaurus provided only a few new terms; it contained no relevant relations. Additionally, as RezoJDM can be enhanced using crowd-sourcing methods and, in particular, games with a purpose, specific game with a purpose assignments have been given to the JeuxDeMots (contribution interface for RezoJDM) players. Today (June 2017) the domain specific subgraph corresponds to 40K terms (approximately 2.8% of the RezoJDM). The overall adaptation process took about 3 weeks.

In the framework of our experiment, we use *taxonomy* relations, *part-whole* relations, *object-mater* relations and, in some cases, *hyponymy* and *characteristic* relations. Running the experiment involves preprocessing steps (identifying multi-word terms, lemmatization, disambiguation using the *refinement scheme* [14] of our resource), browsing

[16] `http://www.jeuxdemots.org/jdm-about.php`

[17] The process of either inferring new relationships from the terms and relationships already present in the network or sourcing new terms and relationships through user contribution or by automatically identifying and retrieving lexical and semantic relationships from texts or other semantic resources.

[18] `http://iate.europa.eu/switchLang.do?success=mainPage\&lang=fr`

[19] The part of the IATE resource that has been used corresponds to the subject field 6006, French

[20] `http://aims.fao.org/fr/agrovoc`

the graph following path constraints, scoring possible incompatibilities, and finally normalizing scores. Among other approaches using a similar plot, [21] use ConceptNet [16] network for the semantic analysis task.

# 4 Method

In this section, we shall detail the pre-processing steps, describe how we move from plain text to the graph-based representation,detail the graph browsing strategy for nutritional incompatibility detection, detail our the method evaluation and discuss the current results yielded by our system.

## 4.1 Pre-processing

The preprocessing step includes text segmentation which relies on the multi-word lexical entities detection, followed by stop-words removal. The multi-word term detection is done in two steps. First, the recipe titles are cut into n-grams using a dynamic-length sliding window ($2 \leq 4$). Second, the segments are compared to the lexical entries present in RezoJDM which is therefore used as a dictionary. In RezoJDM, the multi-word terms are related to their "parts" by the relation typed *locution* (or "phrase"). This relation is useful for compound terms analysis. Bi-grams and trigrams are the most frequent structures in the domain-specific subgraph as well as in RezoJDM. Indeed, they represent respectively 28% and 16% of the overall set of multi-word terms and expressions (300K units). For our experiment, we often opted for trigrams which turned out to be more informative (i.e. correspond to nodes with a higher out degree). The preprocessed text examples are given in the Table 2.

We used a dump RezoJDM[21] stored as a Neo4j[22]graph database. It is also possible to use special purpose API targeted at real time browsing of RezoJDM, Requeter Rezo[23].

## 4.2 From text to graph

Starting from a sequence $w_1, w_2, ...w_n$ of $n$ terms, we build a "context" namely the lemmatized and disambiguated representation of the recipe title under scope. Such context is the entry point to our knowledge resource.
A context *C* is a sequence of nodes (the text order is preserved) : $C_{w_1,w_2,...w_n} = n_{w_1}, n_{w_2}, ...n_{w_n}$ where the node $n_{w_n}$ is the most precise syntactic and semantic representation of the surface form available in our resource. To obtain such representation we search for the node corresponding to the surface form if it exists[24]. Then, we yield its refinement (usage) if the term is polysemic. The irrelevant refinements are discriminated using a list of key terms that define our domain (i.e. thematic subgraph within a

[21] `http://www.jeuxdemots.org/jdm-about.php`
[22] `https://neo4j.com/docs/java-reference/current/javadocs/`
[23] `https://github.com/lirmm-texte/RequeterRezo`
[24] In the opposite case the acquisition process through external resources like Wikipedia, Web and through crowd-sourcing (Games with a purpose) may be triggered (if an open world hypothesis is favored).

| French (*English*) |
|---|
| 1. Carpe à_la_forestière<br>(*Forest_style carp*)<br>2. Carré d'agneau dans_son_jus<br>(*Rack of lamb in_its_own_jus*)<br>3. Waterzoí de lieu_noir<br>(*Coley waterzooí*)<br>4. Verrines de lentilles au saumon_fumé<br>(*Lentil and smoked_salmon verrines*)<br>5. Gâteau_basque à la confiture de cerises_noires<br>(*Basque_cake with black_cherry jam*) |

**Table 2.** Pre-processed recipe titles. Expressions relevant for the domain (i.e. 1,2) as well as multi-word terms (i.e. 3,4), and phrases(i.e.5) are detected, spaces are replaced by the underscore symbol to facilitate the further processing. Sometimes, two segmentations are possible for the same input: $confiture_de_cerises$ and $cerises_noires$. In such cases, the contrast is favored: $black_cherry$ is more informative than $cherry_jam$ as as the semantics of $jam$ is independent from the semantics of $cherry$ or $cherry_jam$. In our resource, semantics can be explored by looking at the outgoing semantic relationships of $confiture$ compared to that of $confiture_de_cerises$ or at least by looking at the *out* degree of the nodes corresponding to these two segments. In this particular case, the concordance between $cerises$ and $noires$ (both in plural) helps to consolidate the choice guided by semantics.

lexical semantic network). The identification of the refinement is done through the cascade processing of relationships of a node typed *refinement*, *domain*, and *meaning*. The choice between several lemmas (i.e. multiple POS problem) is handled by stating[25] : $\forall a \forall b \forall c \forall x, lemma'(b,a) \wedge lemma''(c,a) \wedge pos(x,a) \wedge pos(x,b) \Rightarrow Lemma(b,a)$. For a term $a$ having multiple lemmas such as $b$ and $c$, we check whether the part of speech (represented as a relationship typed $r_pos$) for $b$ and $c$ is the same as the part of speech for $a$. We choose the lemma with the same part of speech as the term $a$. The context creation function returns for *quiche au thon et aux tomates* (*quiche with tuna and tomatos*) the result *[quiche (préparation culinaire), thon (poisson,chair), tomate légume-fruit]*, respectively:"quiche (preparation), tuna (fish flesh), tomato (fruit-vegetable))".
For each node $n_i \in C$ we explore paths $S = ((n_1 a_r n_2), (n_2 a_r n_3), (n_{m-1} a_r n_m))$. The type of relationships we choose for the graph traversal depends on the local category of the node:

1. If $isa("preparation", n_i)$[26], $r \in \{hypo, part-whole, matter\}$.
   This is the case of mixtures, dishes and other complex ingredients;
2. If $isa("ingredient", n_i)$[27], $r \in \{isa, syn, part-whole, matter, charac\}$.
   It is the case of plain ingredients like *tomato*.

The weight of all relations we traverse must be strictly positive. We traverse the graph testing a range of conditions: relevance to the domain of interest $D_{alim}$ (food domain),

[25] First Order Logic notation
[26] *idem*
[27] *idem*

existence of a path of a certain length ($\leq 2$) and type between the candidate node and and the rest of the context under analysis, co-meronymy relation etc.
To obtain relevant results, two conditions are to be fulfilled. First, there must be a disambiguation strategy and a domain filtering. Second, the similarity has to be handled between the $preparation$ and its hyponyms. Indeed, the exploration of the *part-whole* relations of the network refers to *all the possible* ingredients and constituents of the $preparation$. If the $preparation$ has a conceptual role (i.e. "cake"), the *part-whole* relations analysis will output a lot of noise. In our example, it is important to grasp the absence of *pork (meat)* and the presence of *tuna (fish)* in the *quiche* under scope. Therefore, instead of directly exploring the *part-whole* relations of the $quiche(preparation)$, we rather try to find similar $quiche(preparation)$ hyponyms for the context *C*="quiche (preparation), tuna (fish flesh), tomato (fruit-vegetable))" and yield the typical parts they have in common with the generic $quiche(preparation)$. Our function finds the hyponym which maximizes the similarity score. This score is a normalized Jaccard index over all the positive outgoing *part-whole*, *isa*, *matter* relations.
$J(S_C, S_{C_{hypo}}) = \frac{S_C \cap S_{C_{hypo}}}{S_C \cup S_{C_{hypo}}}$ where $C_{hypo}$ is the hyponym context built on the go.
Different threshold values have been experimented. Empirically, the threshold fixed at 0.30 allows capturing generic similarities such as (for our example) *quiche saumon courgette* ("quiche salmon zucchini") score=0.32, a quiche with some fish and a vegetable. More precise similarity corresponds to higher scores.
Using the described strategy, the irrelevant recipes such as *quiche lorraine* (score=0.20) are efficiently discriminated. Once the relevant hyponym is identified, its *part-whole* neighbors can be grasped. A specific graph traversal strategy is used for the LowSalt diet. It includes exploring the *characteristic* relation type for the $preparation$ and its parts.

### 4.3 From graph to associations

**Algorithm 1** Creation of diet incompatibility scores

**Input** : text *RecipeTitle*, lexical semantic network *RezoJDM*, node *Rgime*
**Output** : Key-value pair *Result* $\leftarrow \varnothing$
**Local variables** : list *LContext*, *LNodes*, *LProcessed*, *LSkip*, *LRelationTypes*, *LResult*, float[]*score*
**Local functions** : makeContext, relatedToContext, lookupIncompatible
/* Initialize general lists */
*LSkip* $\leftarrow$ ListOfNodes ( viter)
/*too general terms i.e. *aliment*("food") */
*LRelationsContexte* $\leftarrow$ {r_lemma, r_raff, r_domain, r_meaning}
*LRelationTypes* $\leftarrow$ {r_isa, r_syn, r_part, r_matter, r_holo,r_hypo}
/* Initialize local lists */
*LContext* $\leftarrow$ makeContext(*RecipeTitle*)
*LNodes*$\leftarrow$ relatedToContext (*LContext*,
$x \in \{ContextNode \xrightarrow{R} x > 0\}$, R $\in$ *LRelationTypes*)
/* incompatibility */
**for all** *Node* $\in$ *LNodes* **do**
  LResult $\leftarrow$ lookupIncompatible (*LNodes*, $x \in \{Node \xrightarrow{R} x > 0\}$,
  R $\in$ *LRelationTypes* && $x \xrightarrow{R_{incompatible}}$ *Diet*), score
  *LProcessed* $\leftarrow$ *LProcessed* $\cdot$ Node
**end for**
/*appending RecipeTitle and its scores*/
Result[RecipeTitle][] $\leftarrow$ LResult
**Return** Result

The incompatibility calculation takes as input the list of diets and the queue $F$ containing terms related to main context. This function looks for an incompatibility path $S_{inc}$ such that $N' \in F \wedge S_{inc} = ((N', r_{type}, N), (N, r_{inc}, N_{DIET})) \wedge type \in (holo|isa|haspart|substance|hypo)$.
The output is a key value pair (*diet*, *score*). The score depends on the distance in the RezoJDM graph and on the relation type between the diet and the incompatible node in the context. Besides the *refinement* relation type[28], it is calculated as follows for the distance $d : \frac{1}{1+d}$. The score for the whole context corresponds to the addition of the individual scores of the context nodes. It is adapted in order to bring it closer to a probability distribution and allow further statistical or predictive processing. Starting from the precited context, our system first obtains a list of nodes linked to the context. Then, after following a traversal strategy, it comes up with a list of probabilistic scores for each part of the context. I.e. (w corresponds to $weight$ and d corresponds to $distance$) :

```
recipe : quiche thon tomate
```

[28] Which participates to the disambiguation process

```
context size 3
[quiche>preparation,thon>poisson,tomate>legume]
processing incompatible for quiche>preparation
r_hypo quiche lorraine d=1 w=76
simcontext [quiche,lorraine]
Jaccard index [quiche,lorraine] 0.20
isSimilar=false
```

Futher similarity processing as described in 4.2., then processing of parts shared by quiche and its hyponyms similar to the context. For each part, the *isa*, *part-whole*, *mater* and *characteristic* relations are further explored

```
r_has_part pte d=1 w=6 *dough*
r_has_part gruyere rape d=1 w=105 *grated Gruyere*
r_has_part oeufs d=1 w=105 *eggs*
r_has_part creme d=1 w=105 *cream*
```
etc.

```
raw incompatibility scores quiche>preparation
Diabetes 1.3, LowLactose 3.0, Halal 0.0,
Kosher 0.3, LowCalories 1.5, LowSalt 0.5,
LowGluten 1.8, Hindu 0.5, LowFat 0.8,
Vegan 0.5, Vegetarian 0.3
```

LowLactose incompatibility detected. LowSalt,LowCalories, Kosher, Hindu, Vegetarian and Vegan incompatibilities suspected.

```
processing incompatible for thon>poisson
r_similar bonite d=1 w=26 *skipjack*
r_object_mater poisson>chair d=1 w=30 *fish (flesh)*
r_carac sal d=1 w=36 *salted*
r_carac cru>171869 d=1 w=1 *raw*
r_carac en conserve d=1 w=35 *canned*
r_carac savoureux>got d=1 w=34 *delicious (taste)*
r_isa łtre vivant d=1 w=82 *living being*
r_isa poisson d=1 w=192 *fish*
```
etc.

```
raw incompatibility scores for thon>poisson *tuna (fish)*
Diabetes 0.0, LowLactose 0.0, Halal 0.0,
Kosher 0.0, LowCalories 0.0, LowSalt 2.0,
LowGluten 0.0, Hindu 3.0, LowFat 0.0,
Vegan 3.0, Vegetarian 3.0
```

Vegetarian and Vegan incompatibility confirmed, LowSalt detected, Hindu incompatibility de-

tected. No incompatibilities have been detected for *tomato*.

The output can be "normalized" in a probabilistic fashion according to the following rule applied to the raw score $s \in L_s$ (raw list of scores) in order to produce a probabilistic score $s_p$ : if $s \geq 0.5$, $s_p \leftarrow 1$, if $s \leq 0.5$, $s_p \leftarrow 0.5$. The range of this new score is restricted to three possible values : compatible (0), uncertain (0.5), and incompatible (1).

```
final probabilistic scores
Diabetes 0.5, LowLactose 1.0, Halal 0.0,
Kosher 0.5, LowCalories 1.0, LowSalt 1.0,
LowGluten 1.0, Hindu 1.0, LowFat 1.0,
Vegan 1.0, Vegetarian 1.0
```

Thus, if we imagine a restaurant scenario, where a client would be informed about the strict incompatibility or compatibility of a dish with his or her nutritional restrictions and alerted about some potential incompatibilities that would need further information from the caterer to be confirmed or denied. Given a list of diets, for some of them, we can only output a probabilistic score. Indeed, in our example, we highly suspect the Diabetes incompatibility and the LowCalories incompatibility but nothing in the semantic environment gives us a full confidence.
The list of terms that are not present in the resource (for example, *sibnekh*, *zwiebelkuchen*) is output by the system. It serves to the further improvement of the RezoJDM graph from external resources. The incompatibility scores are also used for the ever ending learning process as they may form candidate terms for RezoJDM, influence the weight of the existing relations etc.

### 4.4 Evaluation and discussion

The system has been evaluated using a incompatibility annotated corpus of 1 500 recipe titles. The evaluation data has been partially collected using structured document retrieval simultaneously with the raw corpus constitution. Relevant meta-tags (HTML tags present in micro-formats[29]) have been used to pre-build incompatibility scores. The overlap between the different diets (vegan recipes are compatible with a vegetarian diet, low fat recipes are convenient for the low calories diet etc.) has been taken into account. The labels have been adjusted and enriched (LowSalt, Kosher diets) by hand by the authors of the system because the structured meta-tags do not contain such information. The obtained evaluation scores are binary. For now, as our system is under improvement, we estimated that the score returned by the system should be $\geq 0.5$ to be ranked as acceptable. Later, a finer grained evaluation will be adopted.
Our results are listed in the Table 3. It is expressed in terms of precision, recall and

[29] Micro-formats($\mu F$) refer to standardized semantic markup of web-pages.

f-measure (F1 score)[30]. The most important score is the precision as, for an allergic or intolerant restaurant customer, even a very small quantity of a forbidden product may be dangerous. The average value correspond to the macro-average[31].

| | Corpus | | | Evaluation set | | |
|---|---|---|---|---|---|---|
| **Diet and scores** | **0** | **0.5** | **1** | **Precision** | **Recall** | **F-score** |
| diabetes | 344 | 1667 | 2989 | 92% | 92% | 92% |
| low lactose | 1410 | 1856 | 1733 | 71% | 73% | 72% |
| halal | 1629 | 3140 | 231 | 65% | 75% | 70% |
| kosher | 1307 | 3453 | 240 | 67% | 60% | 63% |
| low calories | 2540 | 122 | 2338 | 60% | 75% | 67% |
| low salt | 3491 | 1312 | 197 | 88% | 65% | 75% |
| low gluten | 589 | 2961 | 1450 | 80% | 73% | 76% |
| hindu | 568 | 3939 | 493 | 86% | 80% | 83% |
| low fat | 2161 | 2636 | 203 | 67% | 70% | 68% |
| vegan | 454 | 4261 | 285 | 80% | 90% | 85% |
| vegetarian | 360 | 1655 | 2985 | 83% | 90% | 86% |
| **macro-average** | | | | **76%** | **77%** | **76%** |
| **totals** | **14 852** | **27 004** | **13 144** | | | |

**Table 3.** Corpus-based evaluation. We specify the repartition between the 3 possible probabilistic scores in the whole corpus and detail the quality of our scoring for an sub-corpus annotated for evaluation. The evaluation set is a subset of our corpus annotated in incompatibilities. We totally ignore the annotations during the processing.

For the halal diet, there are very few terms in the graph that point to this diet. The LowCalories and Vegetarian diets are well known among the RezoJDM community and well represented in the graph. In 13% of cases, the graph traversal did not return any result as the terms corresponding to the words of the context do not exist in the graph. It was the case of borrowings (such as *quinotto*) or lexical creations[32] (i.e. *anti-tiramisu* as our analysis scheme doesn't take into account morphology). The average number of incompatibility scores $\neq 0$ per recipe title is of about 3.8. Traditional diets (such as halal

[30] F1 score is calculated as follows

$$F_1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall}$$

where *Precision* corresponds to all relevant answers returned by the search and *Recall* to all relevant documents that are successfully retrieved.

[31] Macro average refers to the arithmetic mean, $AM = \frac{1}{n}(a_1 + a_2 + ... + a_n)$, F-score average is therefore the harmonic mean of precision average rate and recall average rate

[32] The term *lexical creation* refers to the neology (creation of new terms). The commonly established general typology of neology distinguishes *denominative* and *expressive* neology. Denominative neology is used to refer to the creation of new lexical units to denominate new concepts, objects, and realities. The expressive neology refers to the use of lexical creation to introduce different subjective nuances.

diet and kosher diet) have been very challenging as the nature of nutritional restriction for them may concern food associations. The low salt diet incompatibility detection is still difficult because salt is everywhere and it is sometimes difficult to find a criterion to separate the property from the possibility of being salted.
Given the specificity of our resource, a low score may depend on the lack of some relevant information in the graph. Thus, the output may be considered as correct (at least at the development stage) but with a low "confidence" which corresponds to a lower mean[33] value.
The mean and standard deviation values reveal this confidence issue (Table 4). A lower mean value together with a low deviation indicates that there were too many uncertain scores ($\leq 0.5$) among the resulting scores of a particular diet. To maximize the confidence related to the score, the knowledge resource must be continuously enhanced.

| **Diet** | **Mean** ($M$) | **Standard Deviation** ($SD$) | **Comment** |
|---|---|---|---|
| diabetes | 0.785 | 0.295 | $M \geq 0.5$ and $SD \leq 0.5 \Longrightarrow$ confidence about the incompatibility |
| low lactose | 0.200 | 0.215 | $M \leq 0.5$ and $SD \leq 0.5 \Longrightarrow$ uniform scores but low confidence |
| halal | 0.240 | 0.239 | $M \leq 0.5$ and $SD \leq 0.5 \Longrightarrow$ uniform scores but low confidence |
| kosher | 0.275 | 0.243 | $M \leq 0.5$ and $SD \leq 0.5 \Longrightarrow$ uniform scores but low confidence |
| low calories | 0.487 | 0.484 | $M$ and $SD$ close to $0.5 \Longrightarrow$ confident but sparse (scores may be very low or quite high), the data is unequally distributed in the resource, more relationship types should be explored |
| low salt | 0.031 | 0.094 | very few incompatibilities detected with low confidence, finer grained approach is needed (i.e. relationship annotation) |
| low gluten | 0.549 | 0.349 | good confidence but scores vary a lot, knowledge consolidation needed |
| hindu | 0.447 | 0.244 | $M$ close to $0.5$ and $SD \leq 0.5 \Longrightarrow$ uniform scores, satisfactory confidence at the consolidation stage |
| low fat | 0.380 | 0.202 | $M \leq 0.5$ and $SD \leq 0.5 \Longrightarrow$ uncertain scores |
| vegan | 0.406 | 0.219 | $M \leq 0.5$ and $SD \leq 0.5 \Longrightarrow$ uncertain scores |
| vegetarian | 0.809 | 0.253 | $M \geq 0.5$ and $SD \leq 0.5 \Longrightarrow$ confidence about the incompatibility, benchmark result |
| **overall** | **0.406** | 0.219 | low confidence but uniform scores |

**Table 4.** Mean values per diet (evaluation)

[33] The mean, as it is understood here, weights each score $s_i$ according to its probability given the dataset, $s_i$. Thus, $\mu$=$\sum s_i p_i$

Today this is achieved using two main strategies: exogenous and endogenous. The first one may include term extraction from corpora, relationship identification based on word embeddings, direct contribution, specific games with a purpose assignments. The second one refers to propagation of the relations already existing in the graph using inference schemes based on the transitivity of the *isa*, *part-whole*, and *synonym* semantic relations as proposed by [24].

### 4.5 Discussion

Using a lexical semantic network for text processing offers some clear advantages. First, the background knowledge management is handy as there is no need to use multiple data structures (lists, vocabularies, taxonomies, ontologies) during the processing. Second,the graph structure supports encoding various kinds of information that can be useful for text processing tasks. Every piece of information encoded using the graph formalism is machine interpretable and can be accessed during the traversal. Thus, the graph interpretation algorithms can be more accessible to domain specialists. Finally, the graph based analysis strategy has an explanatory feature, it is always possible to know why the system returned some particular output. However, two main issues must be tackled : the knowledge resource **population** method and the **filtering** strategy while traversing the resource. Our resource is only available for French. To test the **interoperability of the system**, we run our approach using the ConceptNet common knowledge network and tested it on a restricted set of recipe titles in English that are also available in the French culinary tradition (here we cite the examples of *veal blanquette* and *banana bread*). Then, the obtained probabilistic incompatibility scores have been compared across languages. The values are given as follows: ”score for French [score for English]”.

”blanquette de veau” `["veal blanquette"]`
`Vegan 1.0[0.5],Vegetarian 1.0[0.5]`
`Hindu 1.0[0.5],Diabetes 0.6[0.0],`
`LowLactose 0.5[0.0],Kosher 0.4[0.0],`
`LowFat 0.4[0.0], LowSalt 0.4[0.0]`

”cake à la banane” `["banana bread"]`
`Diabetes 1.0[1.0], LowLactose 0.5[0.0]`
`LowGluten 0.5[0.0], LowFat 0.5[0.0]`
`Vegan 0.5[0.5]`

We can see that the some incompatibilities have been detected using subsumption and hierarchical relations To grasp the *kosher* incompatibility (meat cooked into a creamy sauce) or the LowLactose incompatibility (banana bread may contain some butter), we would need to either extend or specify the *part-whole* relation in order to cover the *matter/substance* relation. We probably need finer grained relations such as *patient*, *receivesAction* etc. This additional testing showed that the approach introduced here is a generic approach, further performance improvement mainly depends on the data contained in the resource.

## 5 Conclusion

We introduced the use of a lexical-semantic network for nutritional incompatibility detection based on recipe titles. The scores obtained by our system can be used for building specific resources for machine learning tasks such as classifier training. Our experiments showed the implementation of a graph browsing strategy for incompatibility detection. Knowledge resource projection on raw text is a relevant method for short text analysis. Among the perspectives of the approach, we can list:

- exogenous approaches (using external resources and processes to populate the graph) for the graph population;
- endogenous approaches to the graph population (i.e. propagating the relevant relationships over the RezoJDM graph, relationship annotation);
- mapping the resource to other existing semantic resources (domain specific ontologies, other wide-coverage semantic resources such as BabelNet[19]);
- making the system evolve towards a multilingual (language independent?) nutritional incompatibility detection.

An important advantage of the system over purely statistical approaches is its explanatory feature, we can always know how the system came to its decision and thus can constantly improve it. The limitations of our contribution are linked to its improvement model which is based on contribution work and the necessity to (weakly) validate the new relations.

## References


1. Ahn, Y.Y., Ahnert, S.E., Bagrow, J.P., Barabsi, A.L.: Flavor network and the principles of food pairing. CoRR abs/1111.6074 (2011)
2. Akkoç, E., Cicekli, N.K.: Semanticook: A Web Application for Nutrition Consultancy for Diabetics, pp. 215–224. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
3. Allen, J.F.: An interval-based representation of temporal knowledge. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence, IJCAI '81, Vancouver, BC, Canada, August 24-28, 1981. pp. 221–226 (1981), http://ijcai.org/Proceedings/81-1/Papers/045.pdf
4. Beltagy, I., Erk, K., Mooney, R.: Semantic parsing using distributional semantics and probabilistic logic. In: Proceedings of ACL 2014 Workshop on Semantic Parsing (SP-2014). pp. 7–11. Baltimore, MD (June 2014), http://www.cs.utexas.edu/users/ai-lab/pub-view.php?PubID=127440
5. Bos, J.: Open-domain semantic parsing with boxer. In: Proceedings of the 20th Nordic Conference of Computational Linguistics, NODALIDA 2015, May 11-13, 2015, Institute of the Lithuanian Language, Vilnius, Lithuania. pp. 301–304 (2015), http://aclweb.org/anthology/W/W15/W15-1841.pdf
6. Chatzikyriakidis, S., Lafourcade, M., Ramadier, L., Zarrouk, M.: Type Theories and Lexical Networks: Using Serious Games as the Basis for Multi-Sorted Typed Systems. In: ESSLLI: European Summer School in Logic, Language and Information. Barcelona, Spain (Aug 2015), https://hal.archives-ouvertes.fr/hal-01216589
7. Cordier, A., Dufour-Lussier, V., Lieber, J., Nauer, E., Badra, F., Cojan, J., Gaillard, E., Infante-Blanco, L., Molli, P., Napoli, A., Skaf-Molli, H.: Taaable: a Case-Based System for personalized Cooking. In: Montani, S., Jain, L.C. (eds.) Successful Case-based Reasoning Applications-2, Studies in Computational Intelligence, vol. 494, pp. 121–162. Springer (Jan 2014), https://hal.inria.fr/hal-00912767
8. Dufour-Lussier, V., Ber, F.L., Lieber, J., Meilender, T., Nauer, E.: Semi-automatic annotation process for procedural texts: An application on cooking recipes. CoRR abs/1209.5663 (2012), http://arxiv.org/abs/1209.5663
9. Fellbaum, C. (ed.): WordNet An Electronic Lexical Database. The MIT Press, Cambridge, MA ; London (May 1998), http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=8106
10. Girju, R., Badulescu, A., Moldovan, D.I.: Automatic discovery of part-whole relations. Computational Linguistics 32(1), 83–135 (2006), https://doi.org/10.1162/coli.2006.32.1.83
11. Hua, W., Wang, Z., Wang, H., Zheng, K., Zhou, X.: Short text understanding through lexical-semantic analysis. In: International Conference on Data Engineering (ICDE) (April 2015), https://www.microsoft.com/en-us/research/publication/short-text-understanding-through-lexical-semantic-analysis/
12. Jermsurawong, J., Habash, N.: Predicting the structure of cooking recipes. In: Mrquez, L., Callison-Burch, C., Su, J., Pighin, D., Marton, Y. (eds.) EMNLP. pp. 781–786. The Association for Computational Linguistics (2015)
13. Kiddon, C., Ponnuraj, G., Zettlemoyer, L., Choi, Y.: Mise en place: Unsupervised interpretation of instructional recipes, pp. 982–992. Association for Computational Linguistics (ACL) (2015)
14. Lafourcade, M.: Making people play for Lexical Acquisition with the JeuxDeMots prototype. In: SNLP'07: 7th International Symposium on Natural Language Processing. p. 7. Pattaya, Chonburi, Thailand (Dec 2007), https://hal-lirmm.ccsd.cnrs.fr/lirmm-00200883

15. Lafourcade, M.: Lexicon and semantic analysis of texts - structures, acquisition, computation and games with words. Habilitation à diriger des recherches, Université Montpellier II - Sciences et Techniques du Languedoc (Dec 2011), https://tel.archives-ouvertes.fr/tel-00649851
16. Liu, H., Singh, P.: Conceptnet — a practical commonsense reasoning tool-kit. BT Technology Journal 22(4), 211–226 (2004)
17. Malmaud, J., Wagner, E., Chang, N., Murphy, K.: Cooking with semantics. In: Proceedings of the ACL 2014 Workshop on Semantic Parsing. pp. 33–38. Association for Computational Linguistics, Baltimore, MD (June 2014), http://www.aclweb.org/anthology/W/W14/W14-2407
18. Mori, S., Sasada, T., Yamakata, Y., Yoshino, K.: A machine learning approach to recipe text processing (2012)
19. Navigli, R., Ponzetto, S.P.: BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence 193, 217–250 (2012)
20. Pinter, B., Vassányi, I., Gaál, B., Mák, E., Kozmann, G.: Personalized Nutrition Counseling Expert System, pp. 957–960. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
21. Poria, S., Agarwal, B., Gelbukh, A., Hussain, A., Howard, N.: Dependency-Based Semantic Parsing for Concept-Level Text Analysis, pp. 113–127. Springer Berlin Heidelberg, Berlin, Heidelberg (2014)
22. Ramadier, L.: Indexation and learning of terms and relations from reports of radiology. Theses, Université de Montpellier (Nov 2016), https://hal-lirmm.ccsd.cnrs.fr/tel-01479769
23. Tasse, D., Smith, N.A.: Sour cream:toward semantic processing of recipes. T.R. CMU-LTI-08-005 p. 9 (May 2008)
24. Zarrouk, M., Lafourcade, M., Joubert, A.: Inference and Reconciliation in a Crowdsourced Lexical-Semantic Network. In: CICLING: International Conference on Intelligent Text Processing and Computational Linguistics. No. 14th, Samos, Greece (Mar 2013), https://hal-lirmm.ccsd.cnrs.fr/lirmm-00816230