BioExcom: Detection and categorization of speculative sentences in biomedical literature

Julien Desclés, Motasem Alrahabi, Jean-Pierre Desclés

BioExcom: Detection and categorization of speculative sentences in biomedical literature

Julien Desclés[1,1], Motasem Alrahabi[1] and Jean-Pierre Desclés[1],

[1] LaLIC Université Paris-Sorbonne
Maison de la Recherche
28 rue Serpente, 75006 Paris, France
{julien.descles, motasem.alrahabi}@gmail.com, jean-pierre.descles@paris.sorbonne.fr

Abstract. Biological research papers are replete with speculative sentences. We present the BioExcom rule-based system, which detects speculations in biomedical literature. Furthermore, it enables to distinguish automatically between prior and new speculations in the analyzed paper. BioExcom is based on the Contextual Exploration processing (hierarchical research of linguistic surface markers with the EXCOM computational platform). To accomplish this task, BioExcom uses also specific linguistic resources established by concise semantic analysis performed by a biologist and a linguist. Our work shows that it is possible to detect and categorize speculative sentences without computational deep linguistic analyses. This work could be useful for biologists who are interested by finding new hypothesis in literature.

Keywords: speculation, hypothesis, biology, contextual exploration, categorization, text mining.

1 Introduction

Biological research papers are replete with speculative sentences, which can be also called *hedges* [1]. For a researcher, it is important to recognize all speculative sentences in a paper or about a given topic. Automatic extraction tools for speculative statements from texts constitute an emerging field which attempts to meet this need [2-7]. Indeed, biological literature is currently characterized by an extended on-line access and an exponential growth [8], which are mostly linked with the development of high-throughput methods and computer science technologies. This huge amount of papers constitutes an extraordinary source of biological facts, knowledge and ideas. However, it is very difficult for a single researcher to keep abreast of all developments [9]. To face this challenge, many systems, based on different Natural Language Processing methods, have been built (for reviews, see [10, 11]).

The Contextual Exploration (CE) is a Natural Language Processing method [12], which constitutes an alternative to classical statistical/machine-learning based

technology and to the search for hard-coded linguistic patterns. Indeed, this linguistic method does not perform any preliminary morpho-syntactic analysis and is based on the hierarchical search for linguistic surface markers expressed in regular expressions and declarative rules. It is implemented in a platform, called EXCOM[2], integrating different linguistic resources for different text mining tasks [13, 14].

This paper presents the BioExcom system, which automatically annotates all speculative sentences in biological full text papers by means of the CE processing, EXCOM computational platform and specific linguistic resources established by concise semantic analysis. Furthermore, BioExcom enables to distinguish automatically between prior and new speculations in a biological paper. We argue that these annotations are useful for biologists, regardless of their domains of interest, to evaluate quickly the content and new output of a paper.

2 Task

2.1 Goal

Our goal is to automatically annotate in biological scientific papers all biological speculative sentences (i.e. sentences containing at least one speculative fragment dealing with a biological issue). We consider only sentences with some clear instances of speculative language (the sentence must contain at least one linguistic element expressing speculation). We also want to categorize them into "*prior speculation*" (speculative sentences cited in the paper, but presented as having been proposed previously) and "*new speculation*" (speculative sentences presented for the first time in the paper or not explicitly presented as prior speculation). All the examples presented below are sentences from biological literature, found in approximately seventy papers.

2.2 Definition of biological speculation in articles

According to our analysis, it is possible to contrast schematically two types of statements in a biological paper if we consider their degree of certainty:

- Demonstrated statements: established facts which are accepted by the scientific community or by the authors of the paper. These can be, for example, biological results, data, observations;
- Speculations (non-demonstrated statements): proposals about a biological issue and explicitly presented as not certain in the paper. These can be, for example, working hypotheses, interpretations/explanations of a fact or purely speculative statements (theoretical considerations).

Others types of statements such as deductions, conclusions, argumentation or discussions…, are NOT considered as speculative but as intermediary statements,

[2] For EXploration COntextuelle Multilingue. Find it out at http://www.excom.fr/

because they either present things more or less as certain, or they do not make a proposal.

It is important to mention that biological speculative sentences have been studied by linguists [1-2], and some sets of guidelines for annotation have been proposed [3, 15]. However, we do not completely agree with this definition and consequently we have proposed our own guidelines[3]. We give here several differences between our guidelines and the others (for more details, see [16]).

Thus, contrary to these prior analyses, we are not interested in detecting a sentence expressing a lack of knowledge/open question because it only asks a question about a biological problem without proposing a mechanism. We also do not consider a sentence as a speculation when the author is being circumspect about some of his statements with the expression "*to our knowledge*". Furthermore, a non informative sentence, i.e. mentioning a hypothesis without explaining it at least partially, is not a speculation according to us. We consider also that a deduction ("*These results indicate that…*") is not a speculation. Finally, the guidelines of Medlock [3] do not take into consideration the speculations denied by some authors or facts. Nevertheless, they should be detected because a speculation can firstly be refused by the scientific community and then be demonstrated and accepted.

To summarize, we want especially to extract ideas and proposals about biological issues from papers without taking into account approvals/negations of them, and we consider speculations as a potential source of relevant information for biologists. This view corresponds to the needs of biologists, as can be seen in the SAWN project [17], which offers to users an online repository of hypotheses in Alzheimer's disease research with links to literature.

2.2 Importance of speculative sentences in biological literature

Speculations are crucial in biomedical papers and text mining tools have to take them into account. Biologists are of course more interested by facts or conclusions and they want to distinguish them from the speculations. But at the same time, annotating speculations in texts can be important for a reader in order to reveal the cognitive articulations of a paper especially in case of hypothesis-driven experiments. Biologists are also interested in knowing all hypotheses about one entity or one topic [2, 17], and this can be explained by several reasons.

Since they give meaning to results, speculative sentences sometimes can carry more useful information than factual sentences, which can be fragmented and cryptic. For example, if we consider the following sentences, the data are complex and not easy to interpret for an untrained person. The interpretations (underlined) appear to be clearer and useful for biologists in the sentences (1) and (2):

(1) "*Interestingly, the UDP-glucosyl pyrophosphorylase (UGP) from C. cryptica was not inhibited by 3-P-glycerate or inorganic phosphate, suggesting that the assimilatory glucan is synthesized outside the plastid [104].*"

(2) "*Heterozygous Foxp2 (R552H)-KI mice also showed slight motor impairment (Fig. 2 A and B), except for a small percentage of the population showing low motor*

[3] http://www.bioexcom.net/

activity, and impairment of USV quality such as short-length USVs (Fig. 2D), suggesting that a particular part of the motor system, but not the entire motor system, is shared with the USV neural system."

In addition, speculative sentences will emphasize important data, which can be very useful in data-collection papers (genomic and post-genomic papers, see [18-21]). They can enable the researchers to anticipate future experimental discoveries or highlight mechanisms which have not yet been well demonstrated or are beyond current biological facts. For "theoretical reviews" or theoretical part of research papers (discussion part), speculations can propose other way to envision biological problems and give new experimental ideas [22-23].

2.3 Categorization into prior and new speculation

Despite the usefulness of speculation extraction from biomedical literature, biologists may need a more finely grained schema. Considering their specific use of literature, we propose the categorization into "*new speculation*" and "*prior speculation*".

Knowing the new speculations of a paper can reveal some of the real new output of it and so help to face one important challenge in text mining: deciding if it is worth spending time on reading carefully the paper. The categorization into prior speculation in a paper highlights the emergence of an idea which is taken into consideration by the scientific community and so can also at least partially give an indication about its importance among the huge amount of speculations in the literature.

3 Automatic annotation of speculative sentences by Contextual Exploration processing

3.1 The Contextual Exploration processing

Contextual Exploration (CE) processing is based on the assumption that a morpho-syntactic parsing can be avoided by the contextual analysis of linguistic surface markers, what has the advantage of low computational cost. CE consists of locating discursive expressions used by an author related to a given viewpoint (hypothesis, conclusions, comments, definitions, causal relations, quotations, opinions related to bibliographic references, etc) [12]. This analysis is performed by a linguist eventually with a specialist of a domain, by collecting and categorizing these specific linguistic expressions of a viewpoint.

The linguistic markers of a viewpoint in CE method used for annotating textual segments (which can be a title, a paragraph, a sentence or a clause), are hierarchical: indicators and clues (both expressed into regular expressions). Indicators correspond to linguistic markers (words, discontinuous expressions…), which carry specific information about the studied domain. These linguistic markers can be relatively independent from the authors' style of writing (for instance, "*we present*", "*in*

conclusion", "*our hypothesis is*", "*is responsible for*"). However, sometimes the simple presence of an indicator does not permit an annotation of the textual segment in which it appears, because the discourse value of the indicator can change according to the context. Consequently, a more precise annotation of the text may be required. To this end, contextual exploration rules must be applied in order to locate, in the textual context of the indicator (the same sentence in the case of BioExcom), one or more linguistic clues, allowing either the removal of the semantic indecision or a more precise segment annotation. Hence, according to what is specified in the CE rule, the looking for clues can be performed in the sentence at the right or/and the left of the indicator or even inside the indicator (see some examples below).

It is worth noting that the CE processing is different from a classical rule-based system -which consists of searching for specific patterns in a text- because of: 1) the use of positive and negative clues, 2) the hierarchy between indicators and clues, 3) the use of text structure.

3.2 Computational architecture of the CE engine and overview of text treatment

The overview of our method for automatic extraction of speculative sentences and search for specific speculations is shown in Fig 1. The architecture is based on the EXCOM platform[4] and will not be detailed here, since it has already been described [14].

In order to be annotated, all texts must go through the following steps:

1) Automatic segmentation of sentences: In order to split the text into sentences, we use a set of rules, which are based on disambiguation of typographical signs (for example: period, semicolon, question mark, etc.). The input files for the segmentation module are raw text files in UTF-8 encoding, in a given language, and the output files are in the XML DocBook format for articles.

2) Automatic annotation: The core of the platform architecture (Fig. 1) consists of a CE engine that manipulates the indicators and clues as linguistic markers and CE rules associated for annotating linguistic segments. The annotation processing consists of the search for indicators of a given viewpoint in the segment considered. The identification of one indicator calls the associated CE rules. When the conditions of these rules are satisfied (that is, a systematic search for contextual clues, also expressed by regular expressions, in the segment), the CE engine attributes the corresponding annotation (organized in a semantic map) to the segment. In addition, the CE engine is able to establish a hierarchy between rules so as to take into account the fact that some indicators or some rules are more indicative than others. Thus, a sentence will be first analyzed according to a first group of rules, then a second one, and so on. This can also prevent, partly, multiple (possibly contradictory) annotations of sentences.

3) Storage of annotations in the base of annotation: Once the texts are annotated, all speculative sentences of the corpus are stored in a "Base of annotations". The annotation scheme of segments contains the following information: the semantic

[4] Principally implemented with Java, JDOM, XML, XLINK, JNLP, etc.

category of the annotation (for example *prior* or *new* speculation), the rule and the indicator used for annotating. The user can choose to consult only *prior* or *new* speculations, and can navigate between speculative sentences and their original context by clicking on the sentence: by doing so, he returns to the original annotated paper.

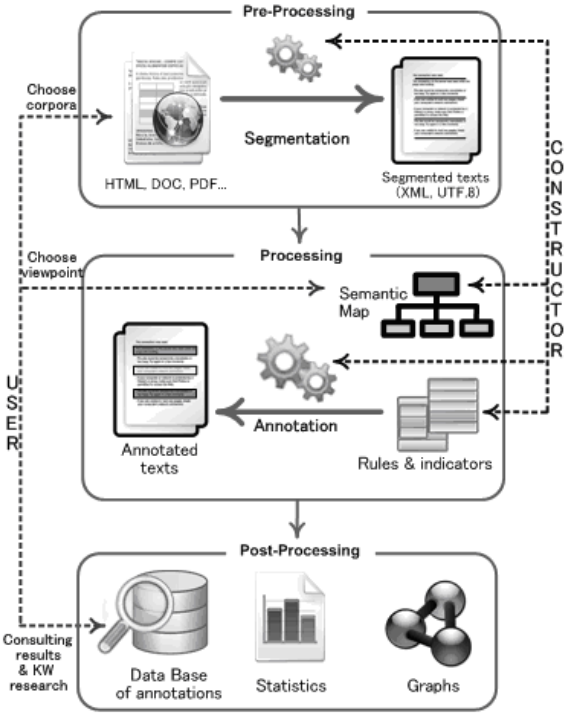

Fig. 1. Overview of the Excom platform.

3.3 The linguistic markers of speculation in biological sentences

The linguistic analysis of speculative sentences by the CE processing consists in studying the linguistic markers of speculation at the sentence level. A careful study, carried out by a biologist and a linguist, on about seventy biological texts (about twenty to built the CE rules and about fifty to test and improve them)[5], has shown that authors use different kinds of specific linguistic markers (or combinations of them) in biological papers, such as:

1) verbs (*to suppose, to suggest, to hypothesize, to propose, to assume*…) ;
2) nouns (*suggestion, hypothesis, speculation*…);
3) adjectives (*convincing, probable, possible, conceivable*…);

[5] From very different biological journals (Nature, Science, Plos biology, PNAS, Plant Physiology, Cell…)

4) adverbs (*possibly, probably, perhaps*…) ;
5) modality verbs (*may, might, could*…) ;
6) conjunction (*if, whether, or*…)

3.4 Categorization of speculative sentences

In order to categorize speculative sentences, we look for some specific verbal aspects (for example the passive present perfect applied to specific verbs for "*prior speculation*"), or the presence of specific constructions (for example, "*we hypothesized*", "*in our theory*" for "*new speculation"*) as indicators. If this kind of markers is not available, we look for the presence or the absence of specific clues, what can highlight the advantage of CE processing for this kind of task.

Here are some of the main others clues used for categorizing speculative sentences:

1) *prior speculation*:

- The presence of bibliographic citations in the sentence, as positive clues at the left or the right of the indicator (sentence 3):
  (3) "*In diatoms, grazing-induced silicification may increase the mechanical resistance of the frustule to copepod mandibules (Hamm et al. 2003).*"
- The presence of others specific words, as positive clues (for example "*recent*" (sentence 4) or "*recent report"* (sentence 5) at the left of the indicator):
  (4) "These recent results with Si and monocots bring not only further support to the theory that Si plays an active role in protecting plants against pathogens, but indicate that this role is not specific to dicots but rather generalized to the plant kingdom."
  (5) "This agrees with a recent report that suggested protein-protein interactions are more conserved within species than across species (49)".

2) *new speculation*:

- The absence of bibliographic citations in the sentence, as negative clues.
- The presence of other specific words, as positive clues (for example "*in this study*" at the left or the right of the indicator in the sentence 6):
  (6) "It is assumed in this study that silicon *layers in* epidermal *cell walls can confer enhanced host resistance to blast."*

3.5 BioExcom implementation

In BioExcom, thirty rules, based on twenty indicator classes (same semantic or grammatical categories), have been built and ranked according to seven priorities. We give here two examples of CE rules in BioExcom. The first one (sentence 7) is the case of the indicator "*could*", which can be either the past form or the conditional form of "*can*". In order to remove this ambiguity, BioExcom checks, in the context of the indicator, the presence or the absence of specific clues expressing conditionality or possibility, such as "*alternatively*" (see the following sentence). Obviously, this method does not allow disambiguating all uses of "*could*", but it correctly recognizes some of them.

(7) "*Alternatively, a soluble Δ9-acyl-ACP desaturase and a membrane-bound Δ9-acyl-lipid desaturase, responsible for the synthesis of 18:1Δ9 and 16:1Δ9, respectively, could co-exist in the plastid of diatoms, similar to the situation found in higher plants.*"

Here is the corresponding simple EC rule, written in a declarative form, used for annotating the sentence as a "*new speculation*":

"could new speculation" CE rule:
Given P a linguistic segment:
If there is in the before-indicator-context a negative clue from the class "bibliographic references"
And If there is in the after-indicator-context a negative clue from the class "bibliographic references"
If there is in the before-indicator context a positive clue from the class "conditionality"
Or If there is in the after-indicator context a positive clue from class "conditionality"
Then : Give the semantic annotation "New Speculation" to P

In order to emphasize the role of negative clues and priority rules, we explain, in the second example (8), how CE processing is able to annotate and to categorize correctly the following sentence.

(8) *"These observations are in agreement with our previous hypothesis suggesting that the particle growth takes place within the voids of the gelatine network so that denser gels will lead to smaller particles."*

Indeed, a classical rule-based system would have to use the specific pattern "*our previous hypothesis suggesting*" to annotate the sentence as a prior speculation, but this pattern is very specific to this sentence. On the contrary, our system applies different rules. It finds in the sentence the indicator "*our () hypothesis suggesting*" (written by a regular expression in the system), but it does not annotate this sentence as a "*new speculation"* because of the presence of "*previous*" as a negative clue inside the indicator space; by using another rule that has less priority than the previous one, the system finds "*hypothesis () that*" as indicator but takes into account "*previous*" as a positive clue for annotating the sentence as "*prior speculation*".

4 Evaluation

4.1 Evaluation methodology

As the concept of speculation in biology is not very clear (see before), we decided to invite experts to give their judgments about results from BioExcom. We devised an evaluation methodology based on the automatic annotation of new and unknown biological articles from different journals by BioExcom and on the random selection of three of them. Between three and five biologists read the version of these papers previously automatically annotated by BioExcom. Before starting the evaluation process, these experts had to read our Annotation guidelines. Then, for each sentence of the annotated articles, they had to say if they were in agreement with the annotation (or the absence of annotation) performed by BioExcom ("*new speculation*" or "*prior speculation*") and, if not, to propose their own annotation according to the categories "*new speculation*", "*prior speculation*", "*undetermined speculation*", "*maybe a speculation but not sure*" and "*not a speculation*".

4.2 Results of the evaluation

The characteristics of the three randomly selected texts in the evaluation process are given in Table 1. The "correct" annotations were determined on the basis of the set of human annotations. These correct annotations are defined as the most frequent or in case of conflict, as the most consensual annotations attributed by the evaluators.

Table 1. Characteristics of the texts used for the evaluation

	Paper 1	Paper 2	Paper 3
Sentences	392	269	375
Words	5536	3396	7001
Duration of annotation (sec)	44,5	31,1	43
Speculative sentences found by BioExcom	30	29	12

The results of the evaluation are given in Table 2. We can note that if we consider also the categorization (*prior* and *new*) of speculative sentences, we observe a weak decrease of performance. Nevertheless, the system finds and categorizes accurately speculative sentences in biological papers. We can make some comments on these results.

First, it has to be mentioned that despite the annotation guidelines and their careful reading by the evaluators, the task of annotating speculative sentences remains quite subjective or difficult (see inter-annotator agreement rates). The sentence in (9) was annotated as speculative by one evaluator, although it was clearly an open question.

Table 2. Results of the evaluation

	Precision	Recall	F-Measure	Inter-annotator agreement
Speculative sentences	98,6	93,0	94,0	78,9
Categorization	89,7	84,6	88,6	67,6

(9) "*At present, however, the genes regulating USV function and the development of the cerebellum and the maturation of Purkinje cells are unknown.*"

In the same way, the sentence in (10) was wrongly annotated by some evaluators as non-speculative, even though it makes a proposal about a biological issue and was correctly annotated by BioExcom thanks to the use of the linguistic markers "*whether*" and "*were not previously established*" by CE processing (see part 3):

(10) "*Which functional domain of Foxp2 or alternative splicing product of Foxp2 functions in the molecular mechanism of mouse USVs and whether the phenotype of Foxp2-KO mice is due to the loss of function of forkhead domain were not previously established.*"

Although the following sentence in (11) was clearly a speculation, it has not been annotated by BioExcom, because of the lack of any specific linguistic marker (indeed "*should*" can not be a marker strong enough to denote accurately a speculation).

(11) "*In contrast, heterozygous Foxp2 (R552H)-KI mice, which showed modest impairment of USVs with different USV qualities and which did not exhibit nuclear aggregates, should provide insights into the common molecular mechanisms between the mouse USV and human speech learning and the relationship between the USV and motor neural systems."*

One other limitation is that some linguistic markers are missing in the implementation of BioExcom. This is the case in the sentence in (12) which has not been recognized as a speculation by the system ("*in principle*" can be a positive clue to disambiguate the indicator "*could*" but was not yet implemented in BioExcom):

(12) "*In principle, the act of transcribing Xist could induce structural changes that could alter chromosome wide function (1).*"

We present here one wrong categorization (but correct extraction as a speculative sentence), which has also been performed by BioExcom: the sentence in (13) was categorized as a "*new speculation*" because of the presence of bibliographic citations.

(13) "*Foxp2 (R552H) nuclear and/or cytoplasmic aggregates caused ER stress in vitro in cell culture (Fig. 5 E–H), probably because of the polyglutamine region, because similar observations were detected in cells expressing polyQ cytoplasmic aggregates (19)*."

The results of this evaluation are very encouraging and indicate that a computationally low cost strategy like CE is efficient for the recognition and categorization of speculations. However, its scale remains quite small. We performed recently another evaluation of BioExcom on a corpus of about 14 500 sentences consisting in a part of the BioScope corpus [15] re-annotated according to our

criteria[6]. This confirmed its ability to detect speculations in full texts by similar results, despite a decrease of recall (Precision = 97,6%; Recall = 77,5%) [17]. Categorization performed by BioExcom has still to be evaluated on a larger scale.

5 Perspectives

The first immediate aim is to finalize the system of automatic annotation in order to offer a user-friendly interface. BioExcom will soon be online to be freely used by researchers to annotate their selected papers.
Another project, which is almost completed and will be published very soon, is to enlarge the system, so as to index all the words of the database of annotated speculative sentences, as it has previously been done with the EXCOM platform. The user (a researcher in biology) may then look for the presence of a list of keywords in the database of speculative sentences. This will enable him to become aware of all hypotheses or speculations proposed about a biological entity (gene or protein for example) or a biological process, which is very useful for researchers [2, 17]. Speculative sentences have the advantage of being very general, whereas the most powerful and useful text mining systems are often very domain-specific (protein phosphorylation [24] for example), probably in order to meet biologists' specific needs [9]. Nevertheless, because of this specificity, it is obvious that, despite their efficiency in recognizing particular patterns in sentences, these systems do not answer entirely the challenge of bridging disjoint literatures. Indeed, most of the time, a knowledge discovery concerns different domains that need to be crossed in a single system in order to establish an unexpected link between two terms [25]. To satisfy this requirement, the user of BioExcom may also look for the combination of two lists of terms (Boolean search) in the speculative sentences, in order to find a hypothesis linking these two terms in very different kinds of biomedical literature. This link may be either not well demonstrated yet, or unknown, but has been already discussed or considered in the literature from a theoretical point of view.
It is also important in the future to connect several dictionaries in order to give the possibility to the researcher to enlarge his list of keywords. In the emergent field of "opinion mining", BioExcom would then be able to better highlight papers describing ideas and proposing hypothesis about precise biological issues, which is a tendency in biological literature and a new challenge for text mining tools [8].

References

1. Hyland K.: The author in the text: Hedging Scientific Writing. Hong Kong papers in linguistics and language teaching 18: 33-42 (1995)
2. Light M, Qiu XY, Srinivasan P.: The Language of Bioscience: Facts, Speculations, and Statements in Between. In HLT-NAACL, ed, Workshop On Linking Biological Literature Ontologies And Databases, pp 17-24 (2004)

[6] available at http://www.bioexcom.net/

3. Medlock B.: Exploring hedge identification in biomedical literature. J Biomed Inform 41: 636-654 (2008)
4. Szarvas G.: Hedge classification in biomedical texts with a weakly supervised selection of keywords, In Proceedings of ACL-08: HLT, pp 281-289, Columbus, Ohio, USA, June 2008 (2008)
5. Kilicoglu H, Bergler S.: Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. BMC Bioinformatics 9 Suppl 11: S10 (2008)
6. Morante R, Daelemans W Learning the scope of hedge cues in biomedical texts, In Proceedings of the Workshop on BioNLP, pp 28-36, Boulder, Colorado, USA, June 2009, ACL (2009)
7. Özgür A, Radev DR.: Detecting Speculations and their Scopes in Scientific Text, In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pp 1398-1407, Singapore, 6-7 August, 2009 (2009)
8. Rebholz-Schuhmann D, Kirsch H, Couto F.: Facts from text--is text mining ready to deliver? PLoS Biol 3: e65 (2005)
9. Cohen KB, Hunter L.: Getting started in text mining. PLoS Comput Biol 4: e20 (2008)
10. Hunter L, Cohen KB.: Biomedical language processing: what's beyond PubMed? Mol Cell 21: 589-594 (2006)
11. Rzhetsky A. SM, Gerstein M.: Seeking a new biology through text mining. Cell 134: 9-13 (2008)
12. Desclés JP, Contextual Exploration Processing for Discourse Automatic Annotations of Texts. In FLAIRS 2006, invited speaker, Melbourne, Florida, pp 281-284 (2006)
13. Djioua B, Flores JG, Blais A, Desclés JP, Guibert G, Jackiewicz A, Le Priol F, Nait-Baha L, Sauzay B.: EXCOM: an automatic annotation engine for semantic information. In FLAIRS 2006, Melbourne, Florida, 11-13 mai, pp 285-290 (2006)
14. Alrahabi M, Desclés JP.: Automatic annotation of direct reported speech in Arabic and French, according to semantic map of enunciative modalities. In 6th International Conference on Natural Language Processing, GoTAL, Gothenburg, Sweden, pp 41-51 (2008)
15. Szarvas G, Vincze V, Farkas R, Csirik J.: The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. In BioNLP ACL-2008 workshop (2008)
16. Desclés J., Makkaoui O., Hacène T.: Detection of speculations in biomedical texts: new perspectives and large-scale evaluation. Proceeding of NeSp-NLP workshop 2010, Sweden, (2010)
17. Clark T and Kinoshita J.: Alzforum and SWAN: The Present and Future of Scientific Web Communities. Briefings in Bioinformatics 8(3):163-171 (2007)
18. Brent R.: Functional genomics: learning to think about gene expression data. Curr Biol 9: R338-341 (1999)
19. Brent R.: Genomic biology. Cell 100: 169-183 (2000)
20. Kell DB, Oliver SG.: Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays 26: 99-105 (2004)
21. Brent R, Lok L.: Cell biology. A fishing buddy for hypothesis generators. Science 308: 504-506 (2005)
22. Bray D.: Reasoning for results. Nature 412: 863 (2001)
23. Blagosklonny MV, Pardee AB.: Conceptual biology: unearthing the gems. Nature 416: 373 (2002)
24. Yuan X, Hu ZZ, Wu HT, Torii M, Narayanaswamy M, Ravikumar KE, Vijay-Shanker K, Wu CH.: An online literature mining tool for protein phosphorylation. Bioinformatics 22: 1668-1669 (2006)
25. Bekhuis T.: Conceptual biology, hypothesis discovery, and text mining: Swanson's legacy. Biomed Digit Libr 3: 2 (2006)