An Ontological Approach for the Quality Assessment of Computer Science Conferences*

Maria Aparecida M. Souto, Mariusa Warpechowski, and José Palazzo M. de Oliveira

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brasil
{souto, mariusa, palazzo}@inf.ufrgs.br

Abstract. Today the proliferation of the availability of the information of scientific events on the Web has created the necessity to offer a quickly access to up-to-date information about the quality of these events. This requirement demands for (semi) automatic tools to speedily provide this information. The human-performed activity of the information quality evaluation is extremely time consuming and easily leads to failures. The application OntoQualis here described was motivated to support the quality evaluation of Scientific Conferences, in the Computer Science area, based on the graduated programs evaluation protocol of the Brazilian agency CAPES. The evaluation mechanism is specified in the QUALIS document specifically designed to assess journals and conferences ranking. This paper presents a brief vision of the ongoing process of domain analysis and ontology prototyping aiming to classify Scientific Conferences: the OntoQualis project. Some results of OntoQualis preliminary evaluation have shown a satisfactory classification level in comparison with CAPES-QUALIS ranking.

1 Introduction

The quality evaluation of scientific productions is mandatory to allocate the constrained resources to an increasing research founding demand. The researchers, research grants demands and graduated programs evaluation work is a real burden to the founding agencies where a high percentage of the total resources is employed in this supporting activity. In some research areas as Biology, Physics and Computer Science the amount of available data in the Web allows an automated quality evaluation process, or at least a semi-automatic process. A single numerical index as the h-index [1] applied to quality assessment may be a dangerous approach if adopted without precautions. In one hand, it is easy to the decision-makers to take a simple measure to support the evaluation process in order to eliminate the complexity of the decision process and the individual decision. On the other hand, it is not fair the use of only one numeric indicator to represent the quality of a researcher's production as a single index measuring only one characteristic of the production. As example we have the

* This work was partially supported by the projects Pronex FAPERGS, grant 0408933; PerXML CNPq. Grant 475.743/2004-0 and CTInfo CNPq, grant 550.845/2005-4. The second and the last authors are partially supported by CNPq.

J.-L. Hainaut et al. (Eds.): ER Workshops 2007, LNCS 4802, pp. 202–212, 2007.

endless discussion about quality versus popularity of a paper. The quality is a multidimensional aggregate variable and for a more complex task as the quality assessment of a scientific conference or a graduate program a full multidimensional evaluation criterion is needed.

The main motivation to the present work on conferences evaluation is the national graduated program evaluation performed triennially by CAPES[1]. The following data was obtained from a public speech of Jorge Guimarães, president of CAPES, Dec 2006: "The Brazilian National System for Graduate Studies encompasses 2.313 programs responsible for 3.624 courses being 2.386 masters and 1.238 doctorates programs. The evaluation of graduate programs is responsibility of the federal agency CAPES and is based in a triennial evaluation performed by academic peers with the participation of 46 evaluation committees and it involves more than 800 evaluators. The triennial evaluation is composed by four processing stages: 1. Data gathering from graduate programs and transmission to CAPES data base; 2. QUALIS Classification of publication media; 3. Definition of evaluation criterion for each area and provide reports for analysis; 4. Data base check and the generation of the evaluation report".

"… The evaluated courses receive grades according to their performance and the following scale: grades 1 and 2 disapprove the program; grade 3 means a regular performance, meeting the minimum demanded; grade 4 is considered a good performance for Masters courses; grade 5 is the maximum grade given to Ms courses; grades 6 and 7 indicate high standard performance with an international level of the course. All the results are public (approved courses and programs, grades, reports) and the evaluation has legal effect. Only the titles and courses approved by the evaluation process have national validity".

The gathering and maintenance of all this data is a huge human effort. The second processing stage the QUALIS generation is a very time-consuming task. QUALIS is a classification of the Journals and Conferences were Brazilian researchers have published papers [2] and is an important referential not only to the graduate programs evaluation but also for the individual researchers evaluation and for grants attribution by the National Research Council – CNPq and other research founding agencies. Presently this work is human-developed and consumes some months to be achieved.

The purpose of the work described in this paper was to develop a classification model to support the (semi) automatic evaluation of Computer Science Conferences (CSC) based on the QUALIS document. This model has generated an ontology prototype and the inference rules to perform the classification. In a first validation test twelve CSCs were evaluated, showing a satisfactory classification level in comparison with CAPES-QUALIS document. We are working in a web-based system that will not only classify a CSC but also presents a dashboard with a multidimensional set of quality indexes.

The paper is organized as follows: Section 2 discusses related work. Section 3 presents the Classification Model for CS Conferences. Section 4 describes the OntoQualis ontology and the classification criteria. Section 5 presents OntoQualis preliminary evaluation, and finally the section 6 presents the conclusion and future works.

[1] CAPES is a Brazilian federal agency charged to promote the human qualification to support the University education.

2 Related Work

For an individual researcher evaluation a single numeric index was recently proposed by J.E. Hirsh, the h-index [1]. This is an interesting numerical evaluator but has some pitfalls. In our opinion the most relevant is the confusion caused by the association of popularity with scientific value. This is a central point in all the quality evaluation processes as the quality of a work must not be considered as the popularity of the work but as the result of a multifaceted evaluation. A superficial paper may be more mediatic than a deep scientific one. This impact-index ranking suffers also a drawback from positive feedback. It is a well-known fact that users access only the first page of a search service answer, this happens in more than 85% of the cases; as a consequence the most cited authors – these appearing in the first page of Scholar Google, as example – are even more cited.

Another bias was recently discovered [3]: the scientific papers being published by bigger and bigger teams, the authors show clearly that the increase in the number of citations is strongly correlated with the number of authors. Exists also a tendency of papers citations of the production of potential reviewers or from colleagues from the same country and others similar bias [4]. These authors have shown in engineering an increasing of 3.72 times in the citations of a paper with more than 5 authors in comparison to a paper of a single author this rate increases to 13.01 times in social sciences for papers published in 2000. Apart these difficulties, the use of an aggregated and easily computable index as the h-index, give an estimate of the importance of the cumulative scientists' production.

For a more complex evaluation, as a scientific conference grading, or the classification of a graduate CS program a wider and multidimensional model of quality must be developed. One alternative for Journals evaluation is the experts' opinion collection and ranking. A recent study [5] presented an extensive analysis of IS Journals. In this survey the perception and impact of IS journals where assessed by an impact factor. An extensive investigation of global community perceptions of IS journals, [6], was employed as reference. One conclusion was: "The data suggests there is consistency in several of the top-rated journals, but wide variety in others". Another analysis from this work is: "Whether perceptions meet with the reality of research dissemination is a question that is open for debate". This analysis supposes that the impact factor is the correct indicator of the dissemination power of a paper; if this assertion is not completely true associated with the above mentioned uncertainties will open a wide research schedule on quality evaluation.

The more extensively employed index, the impact index, alone is clearly subject to some well-founded criticism. We are facing a clear interpretation problem; a consistent experimental protocol must be developed to evaluate the real importance of the index and we need a multi-faceted quality evaluation for the scientific production. This is our research point, how to develop and implement multidimensional metrics to evaluate the scientific production.

3 A Classification Model for CS Conferences

To be able to mange the complexity of the CS Graduated Program assessment we decided to consider initially only the task of classifying CSC. In this work we focused

on the (semi) automation of this task based on the indexes and rules set specified in the CAPES Computer Science area QUALIS document [7] for Conferences evaluation.

According to QUALIS document a CSC could be classified as level "A", "B", and "C". The information is temporally dynamic meaning that the classified CSC could vary annually. The index set defined by the CAPES Committee to classify a CSC is composed by the following indexes: *Impact Index*, *Edition*, *Sponsorship*, *Program Committee*, *Accepted Paper Type*, *Associated Conference*, *Publication* and *Scope*. These indexes are a limited subset of the possible criteria. One of the not considered, in this work, is the acceptance rate, the main criterion was that this index is not widely available and, more important; it is extremely dependent of the conference quality. A poor quality conference attracts a large number of weak papers; in the other hand a top-level conference eliminate the non-competitive ones before the submission. It is clear that some relevant internal data on the reviewing process as the actual number of reviewers per paper is not publicly available for all the conferences. Other interesting issue is an h-index composed from the PC member indexes; we are evaluating some statistical properties of this index. The future inclusion of these indexes and of others will be considered, our expectation is to find the minimal set of indexes allowing a good classification.

The selected indexes were interpreted as follows.

Impact Index – corresponds to that captured from CiteSeer[2] impact index, at the present we are working with evaluation of the h-index[3].
Edition - corresponds to the CSC edition number, this index can influence the classification when combined with others. A well established conference has a higher rating than a first edition one.
Sponsorship – the main CSCs are sponsored by recognized Scientific Institutions. As recognized Institutions we can mention: ACM, IEEE, SIAM, IFIP, W3C and others.
Program Committee – a good CSC has a program committee whose researcher group is composed by recognized researchers assuring serious papers evaluation.
Accepted Paper Type – submitted papers may be classified as: full, short and poster. A full paper has 6 or more pages; short paper are the papers from 3 to 5 pages; and posters the ones having 2 or one pages.
Associated Conference – a Conference may be "Principal" or an "Associated" one. A Workshop is considered as an Associated Conference when it happens in conjunction with Principal Conference. For example, when a workshop is associated to the principal Conference which was classified as level "A", the workshop will be classified as, at last, level "C".
Publication – proceedings published by recognized publisher institutions as ACM, IEEE, SIAM, IFIP, and W3C, as examples.
Scope – the scope index of a Conference is: "Regional", "National", or "International".

The dynamic of these eight criteria according QUALIS document are summarized in Table 1. As we can see in Table 1, conferences level "A" and "B" have two alternatives of classification and conferences level "C" has four. Each alternative corresponds to different sort of combination of the indexes considered. Currently, the *impact index* is considered predominant information to classify a CSC. The alternative to classify a

[2] CiteSeer, http://citeseer.ist.psu.edu/
[3] Publish or Perish software from harzing.com

conference as level "A" take into consideration four indexes: *edition*, *sponsored*, *program committee*, and *accepted paper type*, i.e., the conference *edition* must be greater or equal 4, and the conference must be *sponsored* by recognized institutions in Computer Science area, and it must have a qualified *program committee,* and it have *accepted paper* type "full". *Edition* rule (i.e., >=4) means a traditional conference. Presently, about *program committee*, we are just considering whether the conference has a *program committee*. Nevertheless, colleagues of our team project are working towards a *program committee* quality model.

Another criterion is related to associate conferences. We considered associated conference the workshops which occur together with principal conference. The classification of associated conference depends of classification of the principal conference, i.e., if the principal conference is level "A", the classification of the workshop would be at least level "C".

Table 1. CSC classification criteria summary

	Values	Level "A"		Level "B"		Level "C"			
1. Impact Index	<= 40	x							
	>= 41 and <= 79			x					
	>= 80					x			
2. Edition	<= 3				x				
	>= 4		x						
3. Sponsorship	ACM, IEEE, IFIP, SIAM, W3C, …		x		x				x
4. Program Committee	Yes		x		x			x	
5. Accepted Paper Type	Full		x		x		x	x	
6. Associated Conference	Conference level "A" ➔associated event level "C"						x		
7. Publication	ACM, IEEE, IFIP, SIAM, W3C, …							x	
8. Scope	Regional								x

4 The OntoQualis Prototype

The OntoQualis prototype was developed base on the works of [8, 9]. It implements an ontology based on the classification model for CS Conferences as is shown in Figure 1. The ontology was developed using the plug-in OWL [10] of the Protégé environment [11].

The ontology represents the indexes employed to the conferences classification. The classification task is usual in the context of Information Systems when is required the specification of the pertinence relation within a domain. The use of ontologies in Information Systems generally is oriented to specify and communicate domain knowledge in a generic way and to structure and define the concepts meaning. Besides this, in our work, we also explore the use of the reasoning services aiming to classify CS Conferences in a pre-defined category level.

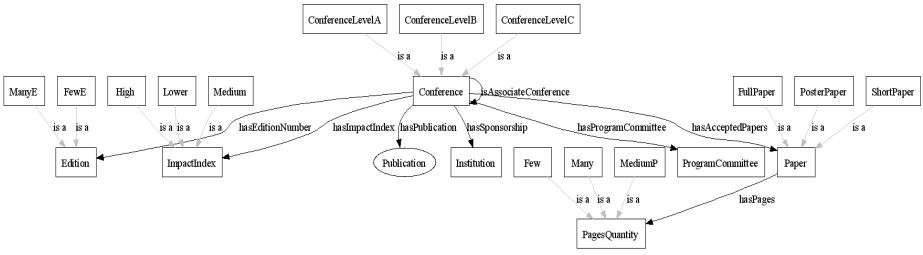

Fig. 1. The OntoQualis Ontology

To accomplish the classification task it is essential to specify which indexes will be measured and how the measurement will be performed. In the following we present the OntoQualis indexes description.

Criterion 1 - ImpactIndex

In the ontological model the Impact Index is represented as the class ImpactIndex and its respective subclasses High, Medium and Lower. The subclass High corresponds to the individuals whose Impact Index value is less or equal 40%; the subclass Medium corresponds to the individuals whose Impact Index value is greater or equal 41% and less or equal 79%; and finally, the subclass Lower corresponds to the individuals whose Impact Index value is greater or equal 80%. In the OntoQualis we have:

ImpactIndex ≡ High ⊓ Medium ⊓ Lower

Criterion 2 – Edition

In the ontological model this index was represented as the class Edition and its respective subclasses FewE and ManyE. The subclass ManyE corresponds to all individual of the domain for which the Edition number value is less or equal 3; and FewE corresponds to that individuals for which this value is greater or equal 4. Thus, in the OntoQualis we have:

Edition ≡ ManyE ⊓ FewE

Criterion 3 – Sponsorship

In the ontological model it represents a property hasSponsorship that relates the Conference class and Institution class. The Institution class can represent three different types of Institutions: Scientific Society, University and Enterprise, which are modeled as subclasses of Institution.

Institution ≡ ScientificInstitutionSociety ⊓ UniversityInstitution ⊓
EnterpriseInstitution

Criterion 4 - Program Committee

In the ontological model it represents the ProgramCommittee class. A Program Committee is composed with chair and members. This is modeled in the ontology through the specification of the properties hasChair and hasMember, which have ProgramCommittee class as domain and Person class as range.

ProgramCommittee ≡ hasChair Person ⊓ hasMember Person

Criterion 5 - Accepted Paper Type
This index was represented in the ontological model as the Papers class and its respective subclasses FullPapers, ShortPapers and PosterPapers. In order to be classified an individual of the domain as pertaining into one of these subclasses it must satisfy some criterion associated with the number of pages of the accepted papers. Then, was created the class PagesQuantity and the respective subclasses Many, MediumP, and Few. The sub-class Many represents the number of pages that is greater or equal 6; the MediumP the number of pages that is less or equal 3 and greater or equal 5; and the Few the number of pages that is less or equal 2.

Papers ≡ FullPapers ⊓ ShortPapers ⊓ PosterPapers

PagesQuantity ≡ Many ⊓ MediumP ⊓ Few

In the Ontology, the papers' types were modeled as the property "hasPages" having the Paper class as Domain and "PagesQuantity" as range, as showed below.

FullPapers ≡ Papers ⊓ hasPages Many

ShortPaper ≡ Papers ⊓ hasPages MediumP

PosterPapers ≡ Papers ⊓ hasPages Few

Criterion 6 - Associated Conference
In the Ontology, all Conferences pertain to the Conference class. In order to distinguish between "Principal" and "Associated" it was defined the "isAssociateConference" property which has Conference class as domain and range.

Criterion 7 – Publication
This index was model as value criterion. To model it in the Ontology it was defined the Publication class and the respective subclasses: BookPublication, ProceedingsPublication and ScientificSocietyPublication, where:

Publication ≡ BookPublication ⊓ ProceedingsPublication ⊓
ScientificSocietyPublication

To represent a recognized Scientific Publisher the "isPrint" property was defined having ScientificSocietyPublication class as domain and the individuals: ACM, IEEE, IFIP, SIAM, W3C, among others, of the ScientificInstitutionSociety class as range. Then we have:

ScientificSocietyPublication ≡ Publication ⊓ isPrint
ScientificInstitutionSociety {ACM IEEE IFIP SIAM W3C}

Criterion 8 – Scope
This index was modeled as value criterion. In the ontology it was represented as the Scope class with the individuals "Regional", "National" and "International":

Scope ≡ {Regional National International}

Conference Classification
As we have mentioned before, our objective is to classify CSC as level "A" or "B" or "C". These levels were modeled in the ontology as subclasses of the class Conference: ConferencelevelA, ConferencelevelB, and ConferencelevelC:

Conference ≡ ConferenceLevelA ⊓ ConferenceLevelB ⊓ ConferenceLevelC

The Conference level classification task involves the evaluation of the eight criteria; each individual of the Conference class is classified as being "A", "B" or "C" considering the evaluation of some combination of these eight criteria. A Conference is classified as level "A" when it has ImpactIndex pertaining to the subclass "High" or when it is sponsored by a recognized Scientific Society and has a Program Committee and has Edition pertaining to the subclass "ManyE" and has Paper pertaining to the subclass "FullPapers". Then, we have:

ConferenceLevelA ≡ (∀ hasImpactIndex High)

ConferenceLevelA ≡ ((∃ hasSponsorship {ACM IEEE IFIP SIAM W3C}) ⊓
(∀ hasProgramCommittee ProgramCommittee) ⊓ (∀ hasEditionNumber ManyE) ⊓
(∃ hasAcceptedPapers FullPapers))

To classify a Conference as level "B" the reasoner must verify if the respective ImpactIndex pertains to the subclass "Medium" or if it is sponsored by a recognized Scientific Society and has a Program Committee and has Edition pertaining to the subclass "FewE" and has Paper pertaining to the subclass "FullPapers".

ConferenceLevelB ≡ (∀ hasImpactIndex Medium)

ConferenceLevelB ≡ ((∃ hasSponsorship {ACM IEEE IFIP SIAM W3C}) ⊓
(∀ hasProgramCommittee ProgramCommittee) ⊓ (∀ hasEditionNumber FewE) ⊓
(∃ hasAcceptedPapers FullPapers))

At last, to classify a Conference as level "C" the reasoner must verify if the respective ImpactIndex pertains to the subclass "Lower" or if the Conference has Paper pertaining to the subclass "FullPapers" and it is associated to the principal Conference previously classified as level "A" or if it has publication pertaining to the subclass "ScientificSocietyPublication" and it has a Program Committee and Paper pertaining to the subclass "FullPapers" or if it is sponsored by a recognized Scientific Society and its scope is "Regional". In order to model this, we have:

ConferenceLevelC ≡ (∀ hasImpactIndex Lower)

ConferenceLevelC ≡ (∃ hasAcceptedPapers FullPapers) ⊓
(∀ isAssociateEvent ConferenceLevelA)

ConferenceLevelC ≡ (∃ hasPublication ScientificSocietyPublication) ⊓
(∀ hasProgramCommittee ProgramCommittee) ⊓ (∃ hasAcceptedPapers FullPapers)

ConferenceLevelC ≡ (∃ hasSponsorship {ACM IEEE IFIP SIAM W3C}) ⊓
(hasScope {Regional})

5 OntoQualis Preliminary Evaluation

The evaluation of the ontological classification model was accomplished in two steps: (i) Conferences data collection and instantiation; (ii) Conferences classification. Our goal was make the comparison between CAPES and OntoQualis CSC classification ranking.

Conferences data collection and instantiation: we chose 12 conferences of Computer Science area, totalizing 154 papers, 840 researchers, 346 Universities, 26 Scientific Societies, 49 enterprises, 12 program committee, and 12 publication titles. Conferences data were collected through web search engines. The information was collected manually by students of a graduated course on Ontology. Presently we are integrating Web collecting tools in an automated environment to populate the Protégé ontology. The collected data corresponds to: *Edition*, *Program Committee* (chair and members), *Publication*, *Sponsorship Institutions*, *Accepted Papers* (title, authors and quantity of pages), *Scope*, and *Impact Index*. The conferences selected were: AH 2006, CAISE 2005, DAMON 2006 (associated with SIGMOD 2006), ER 2005, ICDM 2005, MOBIDE 2006 (associated with SIGMOD 2006), SIGIR 2006, SIGKDD 2005, SIGMOD 2006, UM 2005, VLDB 2005 and WISME 2005 (a standalone conference named workshop).

Conferences classification: In order to perform the conferences classification task, the Racer OWL Reasoner [12] was selected. Through Racer tool, the taxonomy and the ontology consistence were checked. Next, the Racer proceeds with the inference of CSCs level and with the CSCs instantiation into the fitting class (ConferenceLevelA, ConferenceLevelB, ConferenceLelveC).

Table 2. Comparison between the classification results

Scientific Conference	OntoQualis		Qualis-Capes
AH	B	C	A
CAiSE	A	C	A
DAMON	B	C	-
ER	A	C	A
ICDM	B		B
MOBIDE	B		-
SIGIR	A	C	-
SIGKDD	A	B	-
SIGMOD	A	C	-
UM	B	C	A
VLDB	A	C	A
WISME	C		C

It is possible that a conference should satisfy more than one criterion; in this case the reasoner instantiates the conference level in more than one class. As defined in the CAPES document a conference will receive the higher classification level. Table 2 shows the results of the classification generated by OntoQualis and the QUALIS-CAPES official classification which has ranked only the conferences with papers published by Brazilian researchers in the last tree years. The official classification is available on-line[4].

Among these 12 CSCs considered, five were classified as level "A" fulfilling criterion 1 and one was classified as level "A" fulfilling criterion 2. Four CSCs were classified as level "B" fulfilling criterion 3 and one was classified as level "B" fulfilling

[4] http://www.capes.gov.br/avaliacao/webqualis.html
http://qualis.ic.unicamp.br/conferencias/consulta_congressos

criterion 4. Only one CSC was classified as level "C" fulfilling criterion 7. Analyzing these results we can observe that 10 CSCs were classified according official Qualis-Capes rank and only two were not: the AH and UM conferences. We have interpreted these misclassifications as a consequence of the impact index information from Cite-Seer that currently has not been updated.

6 Conclusions and Future Works

The quality evaluation of the scientific work is a complex task. In contrast to other quality evaluation areas as engineering, where some well-defined quantitative measures may be employed, the intellectual production is a multidimensional and subjective task. One of the important quality evaluations of scientific work is associated with Conferences. In this paper we have described our concern towards the development of a (semi) automatic tool taking into consideration several criteria to evaluate the quality of a Conference and, in a near future, alleviate the manual work of the academic community in obtaining this important information.

The first experiments demonstrate that it is possible to reach an (semi) automatic evaluation, compatible with the human-developed classification available from CAPES analysis. In order to obtain an expressive data volume to extensively validate the OntoQualis model, we are working and developing data extraction tools on the Web from the Conferences' homepage.

Up to now, we have worked with the implementation of the QUALIS-CAPES classification model. Nevertheless, we expected to refine the considered indexes and incorporate other ones such as the *program committee* quality index aiming to enrich the present model.

References

1. Hirsh, J.E.: An index to quantify an individual's scientific research output. In: Proceedings of the National Academy of Sciences of the United States of America, November 15, 2005, vol. 102(46), pp. 16569–16572 (2005)
2. CAPES 2007 (March 2007), http://servicos.capes.gov.br/webqualis/
3. Wuchty, S., Jones, B.F., Uzzi, B.: The Increasing Dominance of Teams in Production of Knowledge. Science (published online April 12, 2007), doi:10.1126/science.1136099
4. Leimu, H., Koricheva, J.: What determines the citation frequency of ecological papers? TRENDS in Ecology and Evolution 20(1), 28–32 (2005)
5. Barnes, S.J.: Assessing the value of IS journals. Communications of the ACM 48(1), 110–112 (2005), http://doi.acm.org/10.1145/1039539.1039573
6. Mylonopoulos, V.T.: On site: global perceptions of IS journals. Communications of the ACM 44(9), 29–33 (2001), http://doi.acm.org/10.1145/383694.383701
7. CAPES 2005, http://qualis.ic.unicamp.br/documentos/doc_area_cc-2005-ago05.pdf
8. Noy, N.F., McGuinness, D.L.: Ontology development 101: a guide to creating your first ontology. Technical report KSL-01-05, Stanford Knowledge Systems Laboratory
9. Guarino, N.: Formal ontology and Information Systems. In: FOIS 1998. International conference on Formal Ontologies in Information Systems, Trento, Italy, pp. 3–15 (1998)

10. Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C.: A Practical Guide to Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools (2004), http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf
11. Protegé, julho (2006), http://protege.standford.edu
12. Racer OWL Reasoner, julho (2006), http://www.sts.tu-harburg.de/ r.f.moeller/racer/