A semantic model for scholarly electronic publishing in Biomedical Sciences

Editor(s): Phillip Bourne, University of California at San Diego, USA; Tim Clark, Harvard Medical School and Massachusetts General Hospital, USA; Anita de Waard, Elsevier Laboratories, USA; Alexander Garcia, University of Arkansas for Medical Sciences, USA; Carole Goble, University of Manchester, UK; Steve Pettifer, University of Manchester, UK; David Shotton, University of Oxford, UK
Solicited review(s): Paul Groth, Vrije Universiteit Amsterdam, The Netherlands; Alexander Garcia-Castro, University of Arkansas for Medical Sciences, USA; Tim Clark, Harvard Medical School and Massachusetts General Hospital, USA

Carlos H. Marcondes*, Luciana R. Malheiros and Leonardo C. da Costa
Department of Information Science, Department of Physiology and Pharmacology and Department of Computer Science
University Federal Fluminense, R. Lara Vilela, 126, 24210-590, Niterói, Rio de Janeiro, Brazil, phone 55 21 26299758

Abstract. Despite numerous advancements in information technology, electronic publishing is still based on the print text model. The natural language textual format prevents programs from semantically processing article content. A semantic model for scholarly electronic publishing is proposed, in which the article conclusion is specified by the author and recorded in a machine-understandable format, enabling semantic retrieval, identification of traces of scientific discoveries and knowledge misunderstandings. 89 biomedical articles were analyzed for this purpose. A content model comprising semantic elements and their sequences in articles is developed. Four patterns of reasoning and sequencing of semantic elements were identified in the analyzed articles. The development and testing of a prototype of a Web submission interface to an electronic journal system that partially implements the proposed model are reported.

Keywords: electronic publishing, scientific communication, knowledge representation,ontologies, semantic content processing, scientific discovery, e-Science

* Corresponding author. Email: marcon@vm.uff.br

1. Introduction

Before the advent of the World Wide Web scientific knowledge was distributed across publications in libraries worldwide. The Web is fast becoming a universal platform for the disposal, exchange, and access of knowledge records. An increasing amount of records of human culture—from text, static and motion images, and sound, to multimedia—are now being created directly in digital format.

Since the Philosophical Transactions of the Royal Society in the seventeenth century, scientific articles have become privileged channels of scientific communication. In scientific articles authors bring discoveries into the public knowledge. However scholarly electronic publishing is still based on the print linear text model. These texts are also distributed across various information resources such as digital libraries, electronic journal systems, and repositories. Althoght modern bibliographic information systems exploit the potential of information technology (IT), it is not yet used to directly process the knowledge embedded in the text of scientific articles. Web based publication of scientific articles can serve as *knowledge bases,* as stressed by Gardin [21]. In spite of their digital format, these knowledge bases are comprehensible only to humans. The content of scientific articles deserves critical reading, inquiry, and citation through a long social process until it becomes part of human knowledge. Their textual format hinders the comparison of their semantic content by computers in order to identify gaps and contradictions and agreements in knowledge.

Information Retrieval Systems manage metadata records of scientific articles which provide access to them. Metadata is essential for managing these records in an increasingly complex digital environment. Since the MARC[1] (machine-readable cataloging) record was established in the 1960s, bibliographic record models have slightly changed. A typical bibliographic record comprises sets of database fields, including a flat space of a list of unconnected fields for content description, where keywords or descriptors are assigned, each having an equal weight for retrieval purposes. Content access to documents in modern bibliographic information retrieval systems is still achieved by matching user queries formed by keywords connected by Boolean operators to keywords comprising the bibliographic records, in a manner similar to early bibliographic retrieval and library automation systems.

Typical bibliographic records do not hold explicit relations between elements comprising the content of documents they represent. Boolean operators are too general and lack the semantic expressiveness necessary for content retrieval in specific scientific domains. Relations expressed by Boolean operators are processed as extensive set operations on the keywords included in the bibliographic records, and not as intensive semantic relations between concepts. In a search for policies for dealing with AIDS in PubMed the system recovered an article with the title “A statewide observational assessment of the pedestrian and bicycling environment in hawaii, 2010”, PMID- 22172181, which deals with traffic **policies** including “street accommodations (ie, sidewalks and crossing **aids**)”.

In comparison with the poor expressiveness of the three Boolean operators, the Unified Medical Language System (UMLS) Semantic Network (SN)[2], which is the classification schema of the UMLS National Institutes of Health Metathesaurus, organizes every concept in hierarchy trees, each having as its root a top level Semantic Type. The UMLS SN uses 54 Relation Types to express the semantic relations used between concepts in Semantic Type hierarchies. The UMLS SN holds the permitted relations between Semantic Types. Although this semantically richer schema is supported by the UMLS, the bibliographic record models in databases such as Medline are incapable of exploiting this potential.

Semantic Web (SW) technologies [38] constitute a step forward semantic retrieval and processing in computational environments. The proposal content of a Web document is no longer a matter of keyword match as in conventional computational environments since the 1960s, but instead comprises structured sets of concepts connected by precise meaning relations as in RDF[3] (Resource Description Framework) and RDF Schema[4] statements. Such a rich knowledge representation schema enables software agents to perform “inferences” and more sophisticated tasks based on the document content.

The objetive of this paper is to propose a richer semantic content publishing model, in which scientific claims made by authors throughout articles are expressed by relations between phenomena. In the proposed model, each article, in addition to being published in

[1] MARC Standard, http://www.loc.gov/marc/

[2] UMLS Semantic Network, http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html

[3] RDF, http://www.w3.org/TR/rdf-primer/

[4] RDF Schema Specification, http://www.w3.org/TR/2000/CR-rdf-schema-20000327/

textual format, has its claims also represented as structured relations and recorded in a machine-understandable format using Semantic Web standards such as RDF and OWL [31]. In the proposed model, article records comprise full-text, conventional bibliographic metadata, and semantic metadata conveying the claims made by the author. Based on the processing of these enhanced article records by software agents two research lines were developed: a- the use of such records in semantic retrieval systems; b- comparison of these records with public knowledge—e.g., published scientific articles—or with terminological knowledge bases throughout the Web in order to identify traces of scientific discoveries (this last aspect is reported elsewhere [8], [27]).

The development of these features into software systems may provide scientists with new tools for knowledge retrieval, claim comparison, identification of contradictory claims, use of these claims in different contexts, and identification and validation of new contributions to science made by specific articles.

As other authors previously [3] we propose to engage authors in developing a richer content representation of their own articles; in the approach taken here bibliographic record instances in compliance with the proposed model will be generated by a Web author's submission interface to a journal system, as a byproduct of submitting his/her articles to the system. Such a system, during the upload process of scientific article files, will perform an interactive dialog with authors in order to extract the semantic content of the claims made in the scientific articles and record them in a machine-readable format. The initial steps toward the development of such a system are reported too.

The remainder of this article is organized as follows. The next section presents a review of the theoretical concepts which the proposed publication model is based on along with similar experiences and projects. Section 3 describes the materials and methods used. Section 4 describes the model, its elements and the development of a prototype system of a Web author's submission interface to a journal system which partially implements the model. Finally, section 5 presents the results obtained thus far and discusses the conclusions. It also outlines the future research steps.

2. Related studies

Several alternatives have already been proposed as new types of publications that address the previously discussed issues; some try and exploit SW technologies to enhance scientific communication, management, sharing, and reuse of knowledge; others aim at providing direct access to semantic content of scientific articles. Thus, there is an increasing trend in electronic publishing experiences toward formalizing the text of articles or structuring them, marking them, and identifying significant parts to facilitate more direct reading by humans, potentially by relating the text to formal ontologies [1] as a means to overcome the ambiguity of natural language texts and allow their "semantic" processing by programs.

Scientific articles are documents embedded in definite social relations concerning the scientific communication protocols exhaustively studied in Information Science [5], [6]; with regard to their textual structure, scientific articles are also text-embedded rhetoric/logical theories [2], [11], [24]. The focus of the proposed model is the second aspect, i.e., the reasoning and rhetorical, and the semantic structure of the scientific articles in Biomedical Sciences.

The landscape of the so-called biomedical concept systems has been evolving fast during the last decade. In the literature [30] the term biomedical ontology is a slight imprecise concept naming biomedical concept systems ranging from terminologies used to index scientific literature to highly formal computational ontologies such as OpenGALEN[5]. Their development, evolution, management and integration is a high complex scientific enterprise, as some of them - such as GO - Gene Ontology[6]- were developed recently in order to provide a shared and common terminology to annotate gene products, while others, as MeSH[7] – Medical Subject Headings, which is part of UMLS, has to cope with a legacy of millions of records indexed in bibliographic databases as PubMed and Medline.

Biomedical terminologies are evolving towards knowledge bases [28] as they are becoming formal. According to National Library of Medicine, USA, UMLS Fact Sheet [39]: **"***The purpose of NLM's Unified Medical Language System (UMLS®) is to facilitate the development of computer systems that behave as if they understand the meaning of the language of biomedicine and health***".**

Particularly in Biomedical Sciences, new research methods challenge the conventional Scientific Method and Popperian hypothesis-driven research. The so-called high-throughput methods like DNA microarrays and proteomics [25] allow scientists to process a great amount of data rapidly and in parallel, thus "conducting

[5] OpenGALEN, http://www.opengalen.org
[6] Gene Ontology, http://www.geneontology.org
[7] MeSH, http://www.nlm.nih.gov/mesh/

experiments about which no predictions can be made because no hypotheses have been constructed," as stressed by Westein [23]. This author also stresses the following:

"*Given the layered, evolutionary complexity of biological systems, it will not be possible to understand them comprehensively on the basis of hypothesis-driven research alone. Likewise, it will not be possible to do so solely through "omic" studies of genes, proteins, and other molecules in aggregate. The two modes of research are complementary and synergistic*".

Several alternatives have been considered as new types of publications to address the previously discussed issues and to exploit Semantic Web technologies to enhance scientific communication, management, sharing, and reuse of knowledge, and to provide direct access to the semantic content of scientific articles. The following text comments on these experiences and their conceptual bases.

The Prospect Project[8] is a publishing initiative of the Royal Society of Chemistry, in which terms in the texts of articles that refer to chemical or biological entities have links to dictionaries or ontologies that define them. The Elsevier publishing group is developing a project called Article of the Future associated with the biomedical journal Cell in order to add functionality to several articles, including change in presentation (hierarchical presentations), summary charts, and a section on "Highlights" that briefly outline the conclusions of the article. These facilities are only possible in a Web environment for digitally published articles. Sample articles are available on the project Web site to demonstrate these facilities. A previous study [12] has described the experience of using different semantic technologies in the journal PLoS, including biomedical ontologies, comments on the articles, and an ontology of types or reasons for citation.

HyBrow [35] is a system aimed at helping scientists with hypothesis formulation and evaluation against previous knowledge. The work by Baumgartner and co-authors [41] aimed to identify concepts for extracting protein interaction relations from biomedical text. The approach of [13] to semantic annotations in medical articles considers assertions to be the fundamental units of knowledge. The HypER approach [4] also considers claims to be the basic unit of scientific knowledge. Groth and colleagues [34] present a publication model called nanopublications, consisting of core scientific statements associated with their annotations which specify their context; scientific statements are coded as RDF triples. The Utopia project [37] proposed the assignment of semantic comments to articles in PDF format.

A growing number of scientific publications, especially in the biomedical area, such as the BMJ (British Medical Journal) and the JAMA (Journal of American Medical Association), are using structured abstracts [7] as a way to optimally extract the contents of articles.

3. Materials and methods

The domain of biomedical sciences was chosen because scientific articles in this area follow a strict formal pattern in their texts, with sections defined according to a standard called IMRAD (Introduction, Method, Results, and Discussion) [18].

89 articles in biomedical sciences were analyzed to develop the model with the aim of identifying the semantic elements of scientific methodology, reasoning patterns, and sequencing that combine these elements.

Articles analyzed comprise 3 thematic groups, the last one subdivided in 2 subgroups.

- articles from two outstanding Brazilian research journals, 20 articles from the Memórias do Instituto Oswaldo Cruz, which has its scope mainly in Microbiology, (published during the period 1999-2004), 20 articles from the Brazilian Journal of Medical and Biological Research (published during the period 1998-2004). We used the 20 most downloadable articles of each journal at the moment of the beginning of the research.

- 20 articles about stem cells were also analyzed (published during the period 1994-2004). Stem cells, as an emerging research area in rapid development, were chosen expecting to find articles reporting important discoveries. The articles analyzed were selected from three reviews which present stem cell research development in a historical perspective, pointing out the advances in research, thus of special interest for our work.

- Telomerase 1 group - 15 articles from the Albert Lasker Basic Medical Research Award 2006 key publications were analyzed. This last group is of special interest to the objectives of this research because the articles report, step by step, the rise of a new scientific discovery, the discovery of the telomerase enzyme from 1978 - the first article - to 2001 - the last article of this group. The analysis of this group of articles was guided by an article [14] by the three winners of Lasker Award 2006 which comments the steps toward the discovery of telomerase enzyme.

- Telomerase 2 group - 14 additional articles reporting the development of research on the telomerase enzyme were selected from three

[8] http://www.rsc.org/Publishing/Journals/ProjectProspect/

reviews: the first from Cech [39], a balance of research on telomerase until the date, the previously mentioned paper [14] and articles listed in the timeline available at Telomerase Database site[9]. Each of these 14 articles appears at least in two of the former three reviews.

A comprehensive list of references of the 89 articles, separated by the previous groups is presented in Annex 1.

Each article was analyzed in 4 steps: (1) identify patterns of reasoning developed throughout the article; (2) identify the main conclusion posited by the author in the text; to each article the research group comes to a consensus concerning to the article text which holds the conclusion synthesizing the article's main findings, and to its representation as a relation according to the model proposed; (3) format the claim made in the conclusion as a relation according to the proposed knowledge representation format; and (4) tentatively map each element of the relation to concepts in the UMLS/UMLS SN. Mapping is achieved by comparing terms in the relation extracted in step 3 to MeSH/UMLS terms indexing the article in PubMed records.

The Analysis Form used with examples illustrative of the analysis process for 4 articles are presentd in Annex 2.

A prototype of a submission interface to an electronic journal system was developed, which formats the natural language text of conclusions of articles submitted by authors as semantic relations; this was developed using MetaMap[10], a program that processes biomedical texts to identify terms from the UMLS Thesaurus.

The possibility to apply the model to different areas depends upon the existence of twofold essential elements: the articles must have a clear conclusion that can be stated as a claim and the existence of a terminological databank in machine-readable format which should include terms and relations between them. Biomedical area fulfills both elements. The appliance of the model to different areas needs further experiments.

4. Results and discussion

We have worked for years [9] on the development of a semantic model of electronic publishing. This paper reports the results of a partial implementation of the model as a prototype of a journal submission system. The aim of this model is to achieve a semantically richer content representation of biomedical articles in a program "understandable" format. Such a knowledge representation format allows programs to extract "inferences" about the knowledge content of articles, enabling semantically powerful content retrieval and management relative to current bibliographic Information Retrieval Systems. Instead of manually annotate the text of a paper as in SALT or SWAN [3], [42] the approach taken herein to add semantic metadata to papers is the natural language processing (NLP) of the conclusion of a paper to format it as a semantic relation. Conclusions are typed by authors in addition to conventional bibliographic metadata through a journal submission system.

The proposed model comprises two components: an enhanced semantic record model and a Web interface for authors self-publishing and self-submitting articles to a journal system. The semantic record model *extends* conventional bibliographic record models, which comprise conventional descriptive elements such as authors, title, bibliographic source, and publication date together with content information such as keywords or descriptors. Scientific claims made by authors in their papers are represented as *relations* between two different phenomena or between a phenomenon and its characteristics [17], e.g. "telomere shortening (Phenomenon) causes (Type_of_relation) cellular senescence (Phenomenon)" or "Tetrahymena extracts (Phenomenon) show (Type_of_relation) a specific telomere tranferase activity (Characteristic)". Such relations could be represented as triples of <Antecedent><Type_of_relation><Consequent> .

Our study also includes the development of a prototype system of a Web author's submission interface to a journal system, which implements the model [26], described in Section 4.2, and the use of the general framework proposed to identify discoveries in scientific papers based on two aspects: their rhetoric elements and formats and by comparing the content of the conclusion of articles with terminological data banks [8], [27]. This feature corresponds to step 4 of the analysis process described in section 2 and to the task performed by authors as illustrated in Figure 5.

The following figure shows an overview of the semantic model of electronic publishing, which includes the following components: the Web interface to a system for the submission of articles to electronic publications, the Database, the public Web knowledge base, and the Discoveries identification tool.

[9] http://telomerase.asu.edu/
[10] MetaMap, http://mmtx.nlm.nih.gov/

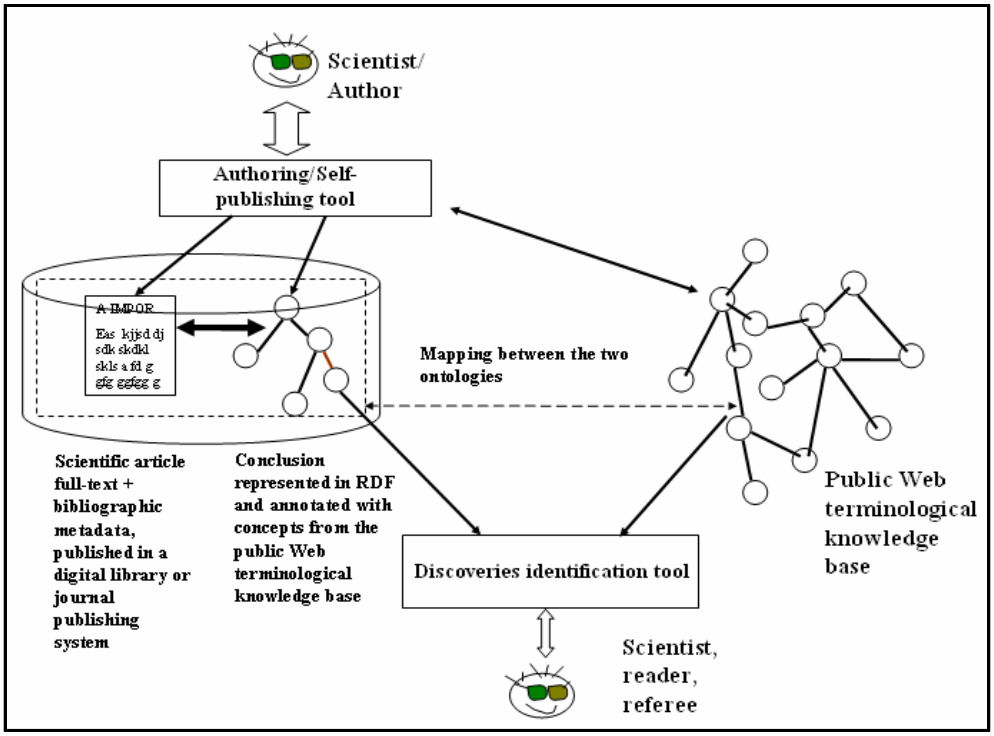

Fig. 1. Overview of the components of the semantic publication model

4.1. A semantic content model for electronic publishing

Relations are the core of the proposed knowledge representation scheme. A relation has the form of an Antecedent (a concept referring to a phenomenon), a Semantic Relation, and a Consequent (a concept referring to a phenomenon or a characteristic of the phenomenon in the Antecedent). A Semantic Relation may be a specific Type_of_relation such as "causes," "affects," or "indicates," or a has/have_characteristic relation. Examples of knowledge representation according to this schema are the following:

- Tetrahymena extracts (Antecedent) show have_Characteristic (Type_of_relation) a specific telomere terminal transferase activity (Consequent).

- Telomere shortening (Antecedent) causes (Type_of_relation) cellular senescence (Consequent).

Relations may also appear in different semantic elements throughout the article text, such as within the Problem that the article addresses, as a *Question*, in which either one of the two *relata* or the type of relation is unknown; in the *Hypothesis;* or in the *Conclusion*. Frequently, the Conclusion also poses new Questions.

Questions, *Hypothesis*, and *Conclusion* are the semantic elements comprising the proposed model. They are the elements related to the knowledge content of an article, which this research aims to identify and record in a machine-processable format. The *Conclusion* is an essential semantic element that synthesizes the knowledge content of an article. In the scope of a recently published article, it is provisional knowledge; however, it is at least guaranteed by the experiment reported in the article. Semantic elements such as *Questions* and *Hypothesis* are important because they enable the evolution of a claim to be determined. Other elements have rhetoric functions, as extensively discussed in [19] and [22], or serve to describe methodological options, the experiment performed, its context, or the obtained results more clearly.

In Biomedical Sciences, there are some standardized methodological procedures, such as PCR (polymerase chain reaction), and some standardized contexts where experiments can take place, for example, in humans (e.g., children, women, embryo), rats, etc.

Thus fhe semantic elements that comprise the proposed record model are the following:

- the problem the article is addressing and the **question** derived from it,

- an **antecedent**,

- a **type_of_relation** (holding the semantic of the relation in a domain, for example, in Biomedical Sciences),

- and the **consequent**.

The **antecedent** and **consequent** may be two different phenomena or a phenomenon and its characteristics.

An empiricall **experiment** is also described with the aim of observing the phenomenon described and specific characteristics of experimental articles divided into:

- **results** – tables, figures, and numeric data reporting the observations made;

- **measure** used;

- a specific **context** where the empirical observations take place, subdivided into:

- **environment** – a hospital, a daycare center, a high school,

- a geographical **place** where the empirical observations take place,

- **time** when the empirical observations occur,

- a specific **population** – pregnant women, early born babies, mice – in which the phenomenon occurs,

- **conclusion** – a set of propositions made by the author as a result of his/her findings.

A **conclusion** corroborates totally or partially the **hypothesis** of an article or negates it. A **conclusion** may also be conclusive or not yet conclusive.

In every analyzed article, concepts found in the antecedent, type_of_relation, and consequent were tentatively mapped (and will be annotated in the future web authoring/publishing tool) to concepts taken from the UMLS. Not all elements are present in all articles.

Articles differ in the way they are built around previously stated hypotheses—those stated by authors other than the author of the current article, or new, original hypotheses, i.e., those stated by the author of the current article. Articles may also differ by the existence of a documented experiment or simply theoretical considerations comparing previously stated hypotheses.

Four patterns of reasoning were found in the analyzed articles, resulting in four article types: *theoretical articles*, which employ abductive (TA) reasoning and *experimental articles,* which may simply be *exploratory* (EE)*,* or employ *inductive* (EI) or *deductive* (ED) reasoning.

Theoretical-abductive (TA) articles analyze different, previous hypotheses, showing their faults and limitations and proposing a new hypothesis; the reasoning is as follows:

*A **problem** is identified, with the following aspects and data…;*

*The **previous hypotheses** (from other authors) are not satisfactory to solve the problem due to the following criticism…;*

*Therefore, we propose this **new hypothesis** (original), which we consider a new pathway to solve the problem.*

Experimental-inductive (EI) articles propose a hypothesis and develop experiments to test and validate it; the reasoning is as follows:

*A **problem** is identified, with the following aspects and data…;*

*A possible solution to this **problem** can be based on the following new **hypothesis…**;*

*We developed an **experiment** to test this **hypothesis** and obtained the following **results**.*

In experimental-inductive articles, a **conclusion** may be mainly one of these alternatives: it corroborates the hypothesis, refutes it, or partially corroborates the hypothesis. However, in some cases, the Conclusion is not one of the former; it simply reports intermediate, and not conclusive, results toward the hypothesis corroboration.

Experimental-deductive (ED) articles use a hypothesis proposed by other researchers cited by the articles' author and apply it to a slightly different context; the reasoning is as follows:

*A **problem** is identified, with the following aspects and data…;*

*In the literature, the **previous hypotheses** (by other authors) have been proposed…;*

*We choose the following **previous hypothesis…**;*

*We enlarge and recontextualize this **hypothesis**; we develop an **experiment** to test it in this new context…;*

*The **experiment** shows the following **results** in this new **context**.*

Experimental-exploratory (EE) articles usually are not hypothesis driven; their objective is to acquire knowledge about a poorly understood scientific phenomenon by performing an **experiment**; the reasoning is as follows:

There is a phenomenon that is poorly understood in a scientific domain.

*We developed an **experiment** that permits the identification of the following characteristics of this phenomenon.*

Within the group of 89 articles that were analyzed, we classified 27 as experimental-inductives (EI), 44 as experimental-deductives (ED), 15 as experimental-exploratories (EE), and 3 as theoretical-abductives (TA).

We are interested not in correlating these patterns with the rhetoric or surface structure of scientifc papers but with the reporting of discoveries in a paper. We found also that these patterns of reasoning are related to fact that

these articles report scientific discoveries. Articles classified as experimental-exploratories (EE) have achieved the lowest grade of mapping of their conclusions to UMLS terms [8], [25].

These basic semantic elements of scientific articles are interrelated and structured. Together with the corresponding bibliographic metadata and article full-text, they form richer article representations in machine-understandable formats and constitute single digital objects that may be stored in a digital library or electronic journal publishing system.

The different reasoning semantic elements and reasoning procedures discussed previously can be formalized in the Model of Knowledge in Articles (MKA), as illustrated in Figure 2 with the hierarchy of classes and componentsproperties.

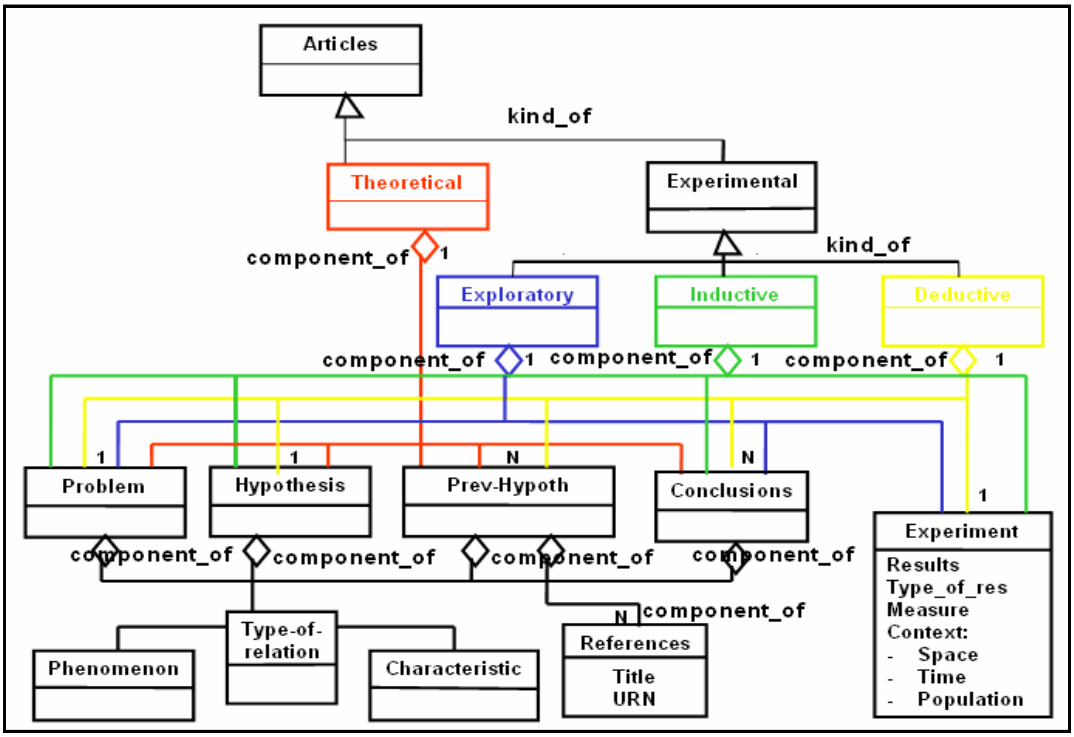

Fig. 2. MKA: model of knowledge representation in articles as a UML class diagram. For clarity different colors show the components of each kind of article

The proposed knowledge representation framework enables the following types of queries to a semantic information retrieval system:

- Which other articles have hypotheses suggesting HPV as the cause of cervical neoplasias in women?

- Which articles have hypotheses suggesting other causes of cervical neoplasias different from HPV in women?

- Which articles have hypotheses suggesting HPV as the cause of cervical neoplasias in groups different from women?

- Which articles have hypotheses suggesting HPV as the cause of pathologies different from neoplasias?

- Which articles have hypotheses suggesting HPV as the cause of cervical neoplasias in different contexts (not in women from the Federal District, Brazil)?

The model also enables queries that may indicate new discoveries, for example, new causes for cellular senescence:

- Which experimental-inductive articles propose (Antecedent?) causes (Type_of_relation) for cellular senescence (Consequent) that are not mapped to UMLS concepts?

- Is there any confirmation of the hypothesis that "Several aspects of both the structural and dynamic properties of telomeres (Antecedent) led to the proposal that telomere replication involves (Type_of_relation) nontemplate addition of telomeric repeats onto the ends of chromosomes (Consequent)?" [20]?

- Who and when first maintained that "the RNA component of telomerase (Antecedent) may be directly involved in (Type_of_relation) recognizing the unique three-dimensional structure of the G-rich telomeric oligonucleotide primers (Consequent)" [10]?

Previous examples, taken from the articles analized, show how the proposed knowledge representation schema may improve semantic retrieval and the use of knowledge in different and unpredicted contexts.

The implementation of the model described in a Web submission interface to an electronic journal system poses the challenges of representing the model, even partially, in a machine-understandable format, and extracting and formatting a relation from the article conclusion. How we addressed these challenges is described as follows. MKA was implemented in RDF as it enables semantic retrieval using SPARQL [36]. The following figure shows the conclusion "telomere replication (Antecedent) involves (Type_of_relation) a terminal transferase-like activity (Consequent)," found in [12], represented in RDF format.

```
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmlns:sa="http://example.org/semarticles/"
   xmlns:umls="http://www.nlm.nih.gov/research/umls/">
   <rdf:Description rdf:about="http://art_id/">
    <dc:title>title</dc:title>
    <dc:creator>creator</dc:creator>
    <dc:subject>subject</dc:subject>
     <dc:date.publisher>date</dc:date.publisher>
     <sa:conclusion>
        <rdf:Description rdf:about="http://art_id/conclusion">
          <sa:antecedent content="telomere
replication">http://www.nlm.nih.gov/research/umls/CUI01</sa:antecedent>
          <sa:type_rel content="involves">http://www.nlm.nih.gov/research/umls/CUI02</sa:type_rel>
          <sa:consequent conten="a terminal transferase-like activity">
http://www.nlm.nih.gov/research/umls/CUI03</sa:consequent>
      </rdf:Description>
     </sa:conclusion>
   </rdf:Description>
</rdf:RDF>
```

Fig. 3. Conclusion of article, represented in RDF. CUI means concept unique identifier.

4.2. Web submission interface to an electronic journal system

A prototype of the submission system was developed to evaluate the dialog with authors and the extraction routine. The prototype of the interface is in its initial phase of development. It is not yet a final version or a production system. It is developed with the specific aim of demonstrating the feasibility of the system proposed. It is developed as a Java application using MetaMap program and Lingpipe[11] library and running in a notebook in order to provide portability to test it with authors interviewed and observe their reactions.

This prototype processes selected parts of the text, namely, the title, abstract, keywords, introduction, methods, and results; the introduction and abstract are used to extract the objective of the article through the identification of phrases such as *objectives of our work…* and *The goal of the present work…* The author is asked by the system to enter the conclusion of the article being submitted. The extraction routine uses a formula, which is based on the frequency of occurrence of a term in the title, abstract, keywords, method, results, and objective, to weigh terms in the conclusion in order to format it from a textual format to a relation. The syntactic components found in the conclusion with higher weights are candidates for the Antecedent and Consequent of the relation.

The natural language processing analysis is performed by the prototype in the following steps:

a - Text in article's Introduction is processed in order to identify which would be the article's Objective; the Objective identified by the system is displayed to authors and he/she is asked to validate the Objective; if he/she do not agree with the identified Objective the system asks the author to enter the Objective.

b - Weka[12] data mining program was previously used to define which sections of an article have strongest correlation between their terms with terms in the Conclusion. Correlations were found in Title, Objective, Abstract, Keyword, Introduction, Material and Methods and Results; sections with strongest correlation were Title, Keywords and Objective.

Text in all these sections was analyzed by MetaMap. MetaMap program parses natural language utterances into components as phrases and assigns to these phrases syntactic categories as VERB_PHRASE (candidates do be relations), NOUN_PHRASE (candidate to be phenomena), PREP_PHRASE (candidate to be context information).

c - NOUN_PHRASEs found in the Conclusion typed by the author are weighted according to their frequency and their presence

[11] Lingpipe, http://alias-i.com/lingpipe/.

[12] Weka, http://www.cs.waikato.ac.nz/ml/weka/

in Title, Objective, Abstract, Keyword, Introduction, Material and Methods and Results;

The weight formula is the following and is based on [16]:

$W(np(n))=A(np(n))+I(np(n))+M(np(n))+R(np(n))+\alpha T(sn(n))+\beta K(sn(n))+\gamma O(sn(n))$,

Where:

- W(np(n)) is the weight of the n th. noun_phrase,
- A(np(n)) is the frequency of the n th. noun_phrase in the Abstract,
- I(np(n)) is the frequency of the n th. noun_phrase in the Introduction section,
- M(np(n)) is the frequency of the n th. noun_phrase in the Material and Method section,
- R(np(n)) is the frequency of the n th. noun_phrase in the Results section,
- T(np(n)) is the frequency of the n th. noun_phrase in the Title,
- K(np(n)) is the frequency of the n th. noun_phrase in the Keywords,
- O(np(n)) is the frequency of the n th. noun_phrase in the Objective.

All sections receive weight 1 except sections with the strongest correlation with the Conclusion as Title, Keywords and Objective which received respectively weights $\alpha=15$, $\beta=5$ and $\gamma=10$. These weights were adjusted comparing results obtained automatically by the system with results obtained by processing articles by the research group.

d - The two first raked not adjacent NOUN_PHRASEs in the Conclusion are the candidates to be the Phenomenon which will be mapped to the two relata of a relation;

e - The system then looks for a VERB-PHRASE occurring in the text between the two NOUN_PHRASEs previously identified as a candidate to be the relation.

A relation dictionary is developed on the base of the 54 relations which comprise UMLS Semantic Network and their synonyms found in WorldNet Dictionary (2010)[13].

f - These linguistic level elements previously identified are then mapped to the article conclusion representation elements, the Phenomenon to Antecedent, the Relation to Type_of_relation, the second Phenomenon to the Consequent. Then the article Conclusion representation elements obtained are mapped to concepts in UMLS. The MetaMap program is used in order to map the article Conclusion representation elements identified to concepts of UMLS Metathesaurus. At this phase we ask the author for validation both of the relation identified and of its mapping to UMLS concepts. Authors will browse UMLS Metathesaurus and will be asked to validate the mapping of the article Conclusion representation identified to the UMLS concepts.

After the author validates the Relation, the system records it as an instance of the MKA according to the format illustrated in Fig. 3, together with the conventional bibliographic metadata and the article full-text. In the future, we plan to integrate this prototype with the PKP Open Journal System[14], an electronic journal system largely used in Brazil. In its present implementation, among the semantic elements that comprise the content model, the prototype processes only the Conclusion.

Some of the steps described above when processing the conclusion "*The results presented herein emphasize the importance to accomplish systematic serological screening during pregnancy in order to prevent the occurrence of elevated number of infants with congenital toxoplasmosis*" [15] are shown in the following Figures.

[13] WordNet, http://wordnet.princeton.edu/

[14] PKP Open Journal System, http://pkp.sfu.ca/

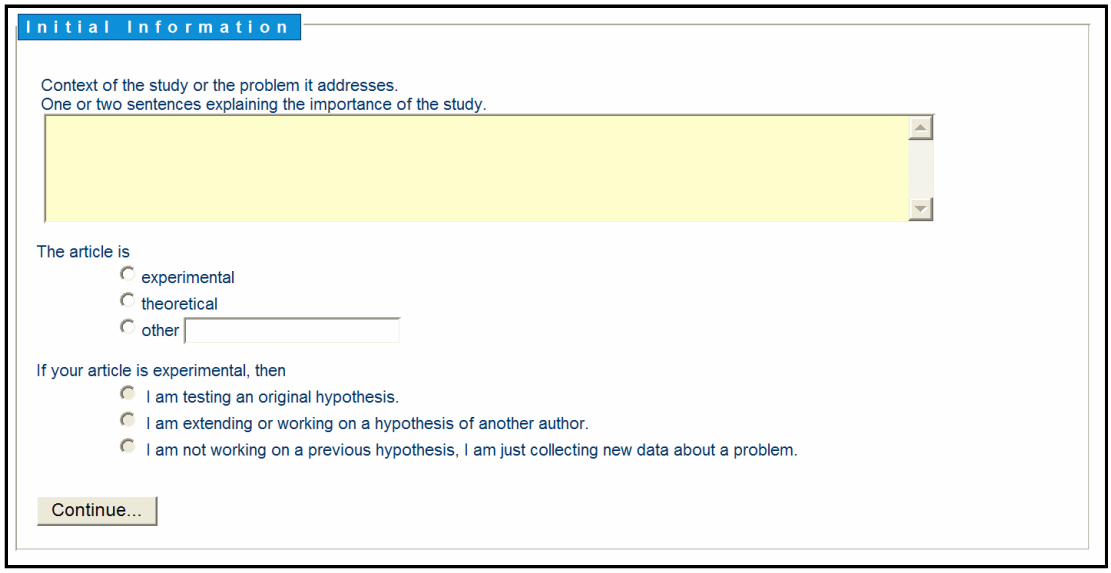

Fig. 4. Author specifies the type of reasoning used in article

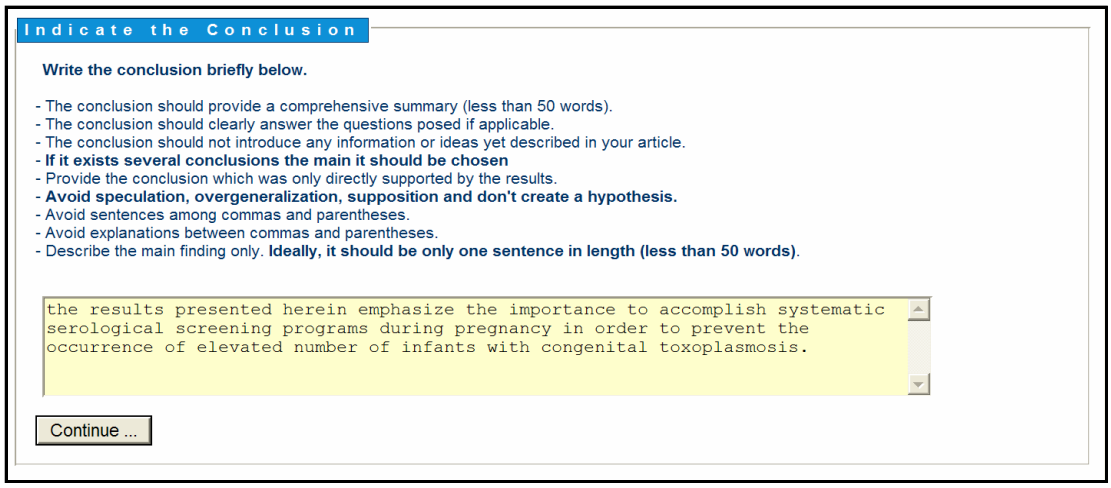

Fig. 5. Author specifies the article conclusion

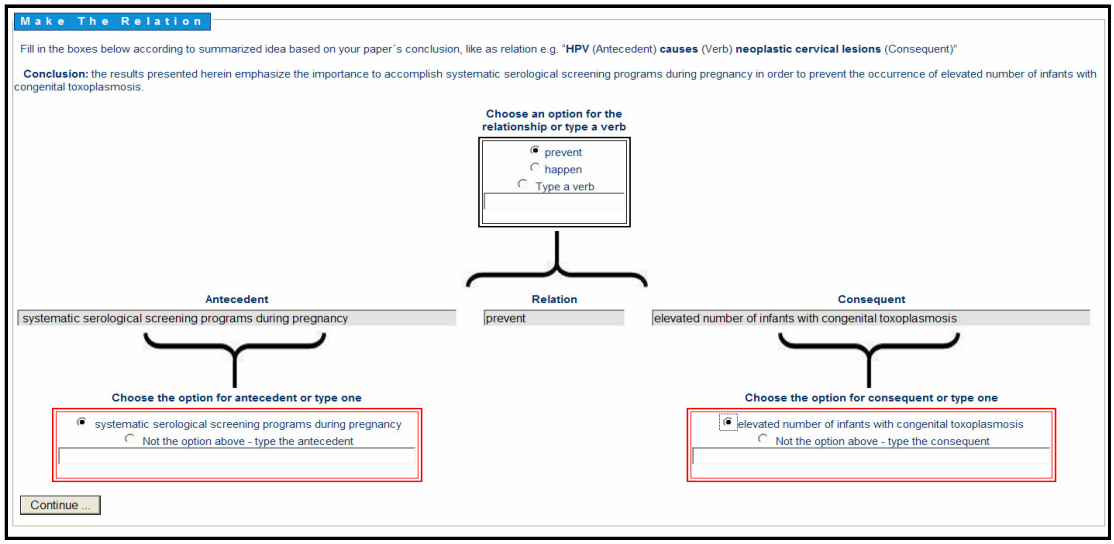

Fig. 6. The article conclusion is formatted as a relation

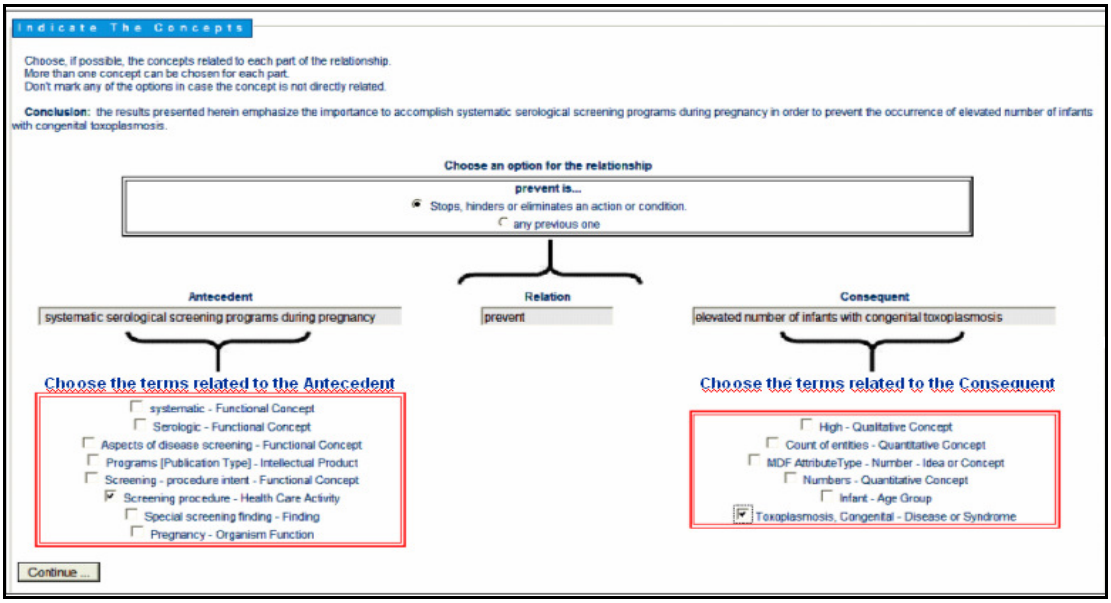

Fig. 7. Authors are asked to map concepts in the article's conclusion to UMLS terms

5. Conclusions

Nowadays, researchers are accustomed to publishing and describing their papers themselves when submitting them to digital libraries, conference management systems, digital repositories or journal systems.

We consider the submission of an article to a journal system to be a privileged process during which *authors are particularly motivated to clarify and disambiguate questions about their articles*. In our proposal, we try to take advantage of this moment. The pathway that seems more feasible to reach this objective is to provide authors with an interactive interface that enables them to validate the automatic NLP carried out by the system. Some elements of the proposed model can be directly obtained by asking questions of the authors, such as whether the article is theoretical or experimental, whether the conclusion confirms or denies the hypotheses, and whether the article is based on the hypothesis of other authors or is original. The NLP of short pieces of text as the articles' conclusions also seems a more feasible and error free option.

A great difficult is to test such a system in operational conditions. This would require the agreement and collaboration of editors and authors. We tested the prototype with six authors of our university and in all cases the system was able to format the conclusion of articles as a relation. Our challenge for now on is to test the system in near operational conditions.

After the claims made by an author from anywhere in the article text, for example, the conclusion, are extracted, they will be represented in a structured form as relations. All these semantic elements can be added to conventional bibliographic elements such as the title, author, abstract, publication data, abstract, and key words, forming richer article representations. This knowledge content will then be represented in a standard machine-understandable format such as RDF. Articles published according to the model proposed can be interlinked and have their content annotated with an increasing number of Web public ontologies, forming a rich knowledge network. This will enable software agents to help scientists to identify and validate new discoveries in Science by comparing the knowledge content of articles with the knowledge content held in public knowledge bases such as the UMLS.

The evidences found of a correlation between articles which conclusion is poorly represented or represented just in a generic level in biomedical terminologies and scientific discoveries are preliminary and deserve more experiments. They may indicate a fruitful pathway to the use of Semantic Web technologies to process articles content.

Although relations play a key role in scientific knowledge, conventional indexing languages do not take them into consideration. The inclusion of relations in knowledge representation makes an expressive difference [43] by enhancing meaning and making more precise the role of subject headings used to represent the document content.

The inclusion of articles conclusions formatted as relations to enhance article metadata is just a proposal. The prototype developed just

aims at testing it feasibility. The full article record lay-out is under development.

The body of scientific literature published on the Web is becoming increasingly vast and complex. It will be necessary for scientists to have enhanced software tools in order to make inferences based on this content. Library and Information Science can go beyond conventional indexing techniques to provide fast access to full-text scientific articles. This would help scientists to directly process the knowledge content of scientific articles and to recover the reasoning that leads to a scientific discovery. The proposed model also points to the standardization of an SkML (Scientific Knowledge Markup Language) encompassing the knowledge content of scientific articles published on the Web, as also proposed by other studies [29], [32], [33]. This opens a new perspective in scientific electronic publishing, knowledge acquisition, storage, processing, and sharing.

This article reports the proposal of a model of semantic publications, not of a production system; such a model aims to evaluate the feasiability of using Semantic Web technologies to enable the processing of articles content by computers. The development of the proposed model to its full potentialities depends on the development of software tools that have not been developed yet. More tests and experiments must be achieved to arrive at full journal and semantic retrieval production systems. Our research group has not been able to fully develop the model to the potentialities outlined here. The proposed model should, however, serve as a starting point that can be discussed and built upon by the scientific community.

Acknowledgments

This research was supported, at different times, by CNPq, CAPES, FAPERJ, and PROPPi/UFF. We would also like to thank Marília Alvarenga Rocha Mendonça.

References

[1] A. H Renear, and C. L. Palmer. Strategic reading, ontologies and the future of scientific publishing. Science 325, pp. 828--832 (2009)

[2] Alan Gross. The Rhetoric of Science. Cambridge, Massachusetts; London: Harvard University Press (1990)

[3]. Anita de Waard. From Proteins to Fairytales: Directions in Semantic Publishing. IEEE Intelligent Systems 25 (2) pp. 83--88 (2010)

[4] Anita de Waard, Simon Buckingham Shum, Annamaria Carusi, Jack Park, Matthias Samwald and Ágnes Sándor. Hypotheses, evidence and relationships: The HypER approach for representing scientific knowledge claims. In: Workshop on Semantic Web Applications in Scientific Discourse, Proc. of the 8th International Semantic Web Conference,. Lecture Notes in Computer Science. Springer Verlag Berlin, Washington DC (2009)

[5] Bernd Frohmann. Documentation redux: Prolegomenon to (another) philosophy of information. Library Trends 52, (3) pp. 387--407 (2004)

[6] Blaise Cronin. Scholarly communication and epistemic cultures. Journal New Review of Academic Librarianship 9, (1) pp. 1--24 (2003)

[7] Carlos Alberto Guimarães. Structured abstracts: Narrative review. Acta Cirúrgica Brasileira, 21, (4) (2006)

[8] Carlos H. Marcondes and Luciana Reis Malheiros. Identifying traces scientific discoveries by comparing the content of articles in biomedical sciences with web ontologies. In: Jaqueline Letha and Binger Larsen (Eds.), Proc. of the 12th ISSI - International Conference on Informetrics and Scientometrics, 2009, Rio de Janeiro, v. 1. pp. 173--177. São Paulo, BIREME/PAHO/WHO, UFRJ (2009)

[9] Carlos H. Marcondes. From scientific communication to public knowledge: the scientific article Web published as a knowledge base. In: Milena Dobreva and Jan Engelen (Eds.), From Author to Reader: Challenges for the Digital Content Chain, Proc. of the 9th ICCC ElPub - International Conference on Electronic Publishing, Leuven, Bélgica, 2005, Leuven, Belgiun pp. 119--127. Peeters Publishing, Leuven (2005)

[10] Carol W. Greider and Elizabeth H.Blackburn. The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell 51, pp. 887--898, (1987)

[11] Charles Bezerman. Shaping written knowledge: Rhetoric of the human sciences. Madison, The University of Wisconsin Press (1988)

[12] David Shotton, Katie Portwin, Graham Klyne and Alistair Miles. Adventures in semantic publishing: Exemplar semantic enhancements of a research article. PLoS Comput. Biol. 5, (4) (2009)

[13] Deendayal Dinakarpandian, Yugyung Lee, Kartik Vishwanath and Rohini Lingambhotla. MachineProse: An ontological framework for scientific assertions. Journal of the American Medical Informatics Association 13, (2) pp. 220--232 (2006)

[14] Elizabeth H Blackburn, Carol W Greider and Jack W Szostak. Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. Nature 12 (10), pp.1133--1138 (2006)

[15] Gesmar Rodrigues Silva Segundo et al. A comparative study of congenital toxoplasmosis between public and private hospitals from Uberlandia, MG/Brasil. Mem. Inst. Oswaldo Cruz 99, (1) (2004)

[16] Harold P. Edmundson. New Methods in Automatic Extracting. Journal of ACM, pp 264--285 (1969)

[17] Ingetraut Dahlberg. Conceptual structures and systematization. International Forum on Information and Documentation 20, (3) pp. 9--24 (1995)

[18] International Committee of Medical Journals Editors, www.icmje.org

[19] J. Skelton. Analysis of the structure of original research papers: an aid to writing original papers for publication. British Journal of General Practice, 44, pp. 455--459 (1994)

[20] Janis Shampay, Jack W. Szostak and Elizabeth H. Blackburn. DNA sequences of telomeres maintained in yeast. Nature 310, pp. 154--157 (1984)

[21] Jean-Claude Gardin. Vers un remodelage des publications savantes: ses rapports avec sciences de l'information. In: Filtrage et Résumé Automatique de l'Information sur les Reseaux - Actes du 3ème Colloque du Chapitre Français de l'ISKO. Paris, Université de Nanterre-Paris X (2001)

[22] John Hutchins. On the structure of scientific texts. In: Proceedings of the 5th. UEA Papers in Linguistics, Norwich pp. 18--39. Norwich, University of East Anglia (1977)

[23] John N. Weinstein. 'Omic' and hypothesis-driven research in the molecular pharmacology of cancer. Current Opinion in Pharmacology 2, (4) pp. 61--65 (2002)

[24] Kevin Ngozi Nwogu. The Medical Research Paper: Structure and Functions. English for Specific Purposes 16, (2) pp. 119--138 (1997)

[25] L. R. Franklin. Exploratory Experiments. In: Proc. of the 19th Biennial Meeting of the Philosophy of Science Assoc. - PSA2004: Contributed Papers, Austin, TX; 2004. Austin, Texas (2004)
[26] Leonardo Cruz da Costa. Um proposta de processo de submissão de artigos científicos à publicações eletrônicas semânticas em Ciências Biomédicas. Tese (doutorado), Programa de Pós-graduação em Ciência da Informação UFF-IBICT. Niterói (2010)
[27] Luciana Reis Malheiros and Carlos Henrique Marcondes. Identificación de los rasgos de descubiertas cientificas en artículos biomedicos. Revista EDICIC, 1 (4), (2011)
[28] Marion G. Ceruti. An Expanded Review of Information-System Terminology. In: Proc. of the AFCEA's Sixteenth Annual Federal Database Colloquium and Exposition, "Information Dominance and Assurance" San Diego, CA, September pp. 21--23 (1999)
[29] M. Hucka, A. Finney, H. Suro, H. Bolouri. System Biology Markup Language (SBML) Level 1: Structures and facilities for basic model definitions. (2003)
[30] Olivier Bodenreider. Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support. In: IMIA Yearbook of Medical Informatics, pp. 67--79 (2008).
[31] OWL Ontology Web Language Overview, http://www.w3.org/TR/owl-features/
[32] P. Murray-Rust, H. S. Rzepa. STMML. A markup language for scientific, technical and medical publishing, Data Science Journal 1, (2), pp. 128--193 (2002)
[33] P. Murray-Rust, H. S. Rzepa. Chemical Markup, XML and the World Wide Web. I: Basic principles, Journal of Chemical Information and Computer Science 39, pp. 928--942 (1999)
[34] Paul Groth, Andrew Gibson and Jan Velterop. The anatomy of a nanopublication. Information Services & Use 30, pp.51--56 (2010)
[35] S. A. Racunas, N. H Shah, I. Albert, and N. V. Fedoroff. HyBrow: a prototype system for computer-aided hypothesis evaluation. Bioinformatics 20, (1) pp. 257--264 (2004)
[36] SPARQL Query Language for RDF (2008), http://www.w3.org/TR/rdf-sparql-query/
[37] Teresa K. Attwood, Douglas B. Kell, Philip Mcdermott, James Marsh, Steve R. Pettifer and David Thorne. Calling international rescue: knowledge lost in literature and data landslide Biochemical Journal, Dec (2009)
[38] Tim Berners-Lee, James Hendler, and Ora Lassila, O. The semantic web, Scientific American. (2001)
[39] Thomas R. Cech. Beginning to understand the end of the chromosome. Cell, 43, pp.405--413 (2004).
[39] Unified Medical Language System Fact Sheet, http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html
[41] William A Baumgartner, Zhiyong Lu, Helen L Johnson, J Gregory Caporaso, Jesse Paquette, Anna Lindemann, Elizabeth K White, Olga Medvedeva, K Bretonnel Cohen and Lawrence Hunter. Concept recognition for extracting protein interaction relations from biomedical text. Genome Biol. 9 (Suppl 2), (2008)
[42] Y. Gao, J. Kinoshita, E. Wu, E. Miller. R. Lee, A. Seaborne, S. Cayzer and T. Clark, SWAM: a distributed knowledge infrastructure for Alzeimer disease research, *Journal of Web Semantic* 4 (3), (2006).
[43] Y Kajikawa; K Abe; S Noda. Filling the gap between researchers studying different materials and different methods: a proposal for structured keywords. Journal of Information Science 32, pp. 511--524 (2006)

Annex 1 – References of the articles analyzed

Memórias do Instituto Oswaldo Cruz

1
CAMARA, Geni NL, CERQUEIRA, Daniela M, OLIVEIRA, Ana PG *et al*. Prevalence of human papillomavirus types in women with pre-neoplastic and neoplastic cervical lesions in the Federal District of Brazil. *Mem. Inst. Oswaldo Cruz*. [online]. Oct. 2003, vol.98, no.7 [cited 10 March 2005], p.879-883.
2
ALVES, Tânia Maria de Almeida, SILVA, Andréia Fonseca, BRANDAO, Mitzi *et al*. Biological screening of Brazilian medicinal plants. *Mem. Inst. Oswaldo Cruz*. [online]. maio/jun. 2000, vol.95, no.3 [citado 18 Agosto 2005], p.367-373.
3
ANDRIGHETTI-FROHNER, CR, ANTONIO, RV, CRECZYNSKI-PASA, TB *et al*. Cytotoxicity and potential antiviral evaluation of violacein produced by Chromobacterium violaceum. *Mem. Inst. Oswaldo Cruz*. [online]. Sept. 2003, vol.98, no.6 [cited 25 September 2005], p.843-848.
4
SMITH, HM, DEKAMINSKY, RG, NIWAS, S *et al*. Prevalence and intensity of infections of Ascaris lumbricoides and Trichuris trichiura and associated socio-demographic variables in four rural Honduran communities. *Mem. Inst. Oswaldo Cruz*. [online]. abr. 2001, vol.96, no.3 [citado 18 Agosto 2005], p.303-314.
5
HOLETZ, Fabíola Barbiéri, PESSINI, Greisiele Lorena, SANCHES, Neviton Rogério *et al*. Screening of some plants used in the Brazilian folk medicine for the treatment of infectious diseases. *Mem. Inst. Oswaldo Cruz*. [online]. out. 2002, vol.97, no.7 [citado 18 Agosto 2005], p.1027-1031.
6
PAIVA, Selma Ribeiro de, FIGUEIREDO, Maria Raquel, ARAGAO, Tânia Verônica *et al*. Antimicrobial activity in vitro of plumbagin isolated from Plumbago species. *Mem. Inst. Oswaldo Cruz*. [online]. Oct. 2003, vol.98, no.7 [cited 23 November 2005], p.959-961.
7
ROSA, Luiz Henrique, MACHADO, Kátia M Gomes, JACOB, Camila Cristina *et al*. Screening of Brazilian basidiomycetes for antimicrobial activity. *Mem. Inst. Oswaldo Cruz*. [online]. out. 2003, vol.98, no.7 [citado 18 Agosto 2005], p.967-974.
8
PINHEIRO, Lucimar, NAKAMURA, Celso Vataru, DIAS FILHO, Benedito Prado *et al*. Antibacterial xanthones from Kielmeyera variabilis mart. (Clusiaceae). *Mem. Inst. Oswaldo Cruz*. [online]. June 2003, vol.98, no.4 [cited 25 September 2005], p.549-552.
9
BETANCUR-GALVIS, LA, SAEZ, J, GRANADOS, H *et al*. Antitumor and Antiviral Activity of Colombian Medicinal Plant Extracts. *Mem. Inst. Oswaldo Cruz*. [online]. July 1999, vol.94, no.4 [cited 30 October 2005], p.531-535.
10
CIRAK, Meltem Yalinay, KALKANCI, Ayse and KUSTIMUR, Semra. Use of molecular methods in identification of Candida Species and evaluation of fluconazole resistance. *Mem. Inst. Oswaldo Cruz*. [online]. Dec. 2003, vol.98, no.8 [cited 15 October 2005], p.1027-1032.
11
NAKAMURA, Celso Vataru, UEDA-NAKAMURA, Tania, BANDO, Erika *et al*. Antibacterial activity of Ocimum gratissimum L. essential oil. *Mem. Inst. Oswaldo Cruz*. [online]. Sept. 1999, vol.94, no.5 [cited 15 October 2005], p.675-678.
12
CAVALCANTI, Eveline Solon Barreira, MORAIS, Selene Maia de, LIMA, Michele Ashley A *et al*. Larvicidal Activity of essential oils from Brazilian plants against Aedes aegypti L. *Mem. Inst. Oswaldo Cruz*. [online]. Aug. 2004, vol.99, no.5 [cited 30 October 2005], p.541-544.
13
MAGALHAES, Aderbal Farias, TOZZI, Ana Maria Goulart de Azevedo, SANTOS, Celira Caparica *et al*. Saponins from Swartzia langsdorffii: biological activities. *Mem. Inst. Oswaldo Cruz*. [online]. July 2003, vol.98, no.5 [cited 23 November 2005], p.713-718.

14
CARVALHO, Ana Fontenele Urano, MELO, Vânia Maria Maciel, CRAVEIRO, Afrânio Aragão *et al*. Larvicidal activity of the essential oil from Lippia sidoides cham. against Aedes aegypti linn. *Mem. Inst. Oswaldo Cruz*. [online]. June 2003, vol.98, no.4 [cited 15 November 2005], p.569-571.
15
SEGUNDO, Gesmar Rodrigues Silva, SILVA, Deise Aparecida Oliveira, MINEO, José Roberto *et al*. A comparative study of congenital toxoplasmosis between public and private hospitals from Uberlândia, MG, Brazil. *Mem. Inst. Oswaldo Cruz*. [online]. Feb. 2004, vol.99, no.1 [cited 05 November 2005], p.13-17.
16
ALMEIDA, Marcos C de, SILVA, Alan C, BARRAL, Aldina *et al*. A simple method for human peripheral bloodmonocyte Isolation. *Mem. Inst. Oswaldo Cruz*. [online]. Mar./Apr. 2000, vol.95, no.2 [cited 05 November2005], p.221-223.
17
ANTAS, Paulo RZ, CARDOSO, Fernando LL, OLIVEIRA, Eliane B *et al*. Whole blood assay to access T cell-immune responses to Mycobacterium tuberculosis antigens in healthy Brazilian individuals. *Mem. Inst. Oswaldo Cruz*. [online]. Feb. 2004, vol.99, no.1 [cited 09 September 2005], p.53-55.
18
ARAUJO, Maria Ilma, HOPPE, Bradford S, MEDEIROS JR, Manoel *et al*. Schistosoma mansoni infection modulates the immune response against allergic and auto-immune diseases. *Mem. Inst. Oswaldo Cruz*. [online]. Aug. 2004, vol.99 suppl.1 [cited 25 September 2005], p.27-32.
19
MARTINEZ-ESPINOSA, Flor Ernestina, DANIEL-RIBEIRO, Cláudio Tadeu and ALECRIM, Wilson Duarte. Malaria during pregnancy in a reference centre from the Brazilian Amazon: unexpected increase in the frequency of Plasmodium falciparum infections. *Mem. Inst. Oswaldo Cruz*. [online]. Feb. 2004, vol.99, no.1 [cited 16 October 2005], p.19-21.
20
WILSON, R Alan, CURWEN, Rachel S, BRASCHI, Simon *et al*. From genomes to vaccines via the proteome. *Mem. Inst. Oswaldo Cruz*. [online]. Aug. 2004, vol.99 suppl.1 [cited 15 October 2005], p.45-50.

Brazilian Journal of Medical and Biological Research

1
COIMBRA, C.G. and JUNQUEIRA, V.B.C. High doses of riboflavin and the elimination of dietary red meat promote the recovery of some motor functions in Parkinson's disease patients. *Braz J Med Biol Res*. [online]. Oct. 2003, vol.36, no.10 [cited 21 January 2006], p.1409-1417.
2
NASCIMENTO, A.L.T.O., VERJOVSKI-ALMEIDA, S., VAN SLUYS, M.A. *et al*. Genome features of Leptospira interrogans serovar Copenhageni. *Braz J Med Biol Res*. [online]. Apr. 2004, vol.37, no.4 [cited 21 January 2006], p.459-477.
3
COVAS, D.T., SIUFI, J.L.C., SILVA, A.R.L. *et al*. Isolation and culture of umbilical vein mesenchymal stem cells. *Braz J Med Biol Res*. [online]. Sept. 2003, vol.36, no.9 [cited 21 January 2006], p.1179-1183.
4
DOMENICE, S., CORREA, R.V., COSTA, E.M.F. *et al*. Mutations in the SRY, DAX1, SF1 and WNT4 genes in Brazilian sex-reversed patients. *Braz J Med Biol Res*. [online]. Jan. 2004, vol.37, no.1 [cited 23 January 2006], p.145-150.
5
LATHA, M. and PARI, L. Effect of an aqueous extract of Scoparia dulcis on blood glucose, plasma insulin and some polyol pathway enzymes in experimental rat diabetes. *Braz J Med Biol Res*. [online]. Apr. 2004, vol.37, no.4 [cited 25 January 2006], p.577-586.
6
JAVORKA, M., ZILA, I., BALHAREK, T. *et al*. Heart rate recovery after exercise: relations to heart rate variability and complexity. *Braz J Med Biol Res*. [online]. Aug. 2002, vol.35, no.8 [cited 25 January 2006], p.991-1000.
7
AVILA, R., BOTTINO, C.M.C., CARVALHO, I.A.M. *et al*. Neuropsychological rehabilitation of memory deficits and activities of daily living in patients with Alzheimer's disease: a pilot study. *Braz J Med Biol Res*. [online]. Nov. 2004, vol.37, no.11 [cited 25 January 2006], p.1721-1729.
8
RAMOS, C.R.R., ABREU, P.A.E., NASCIMENTO, A.L.T.O. *et al*. A high-copy T7 Escherichia coli expression vector for the production of recombinant proteins with a minimal N-terminal His-tagged fusion peptide. *Braz J Med Biol Res*. [online]. Aug. 2004, vol.37, no.8 [cited 25 January 2006], p.1103-1109.
9
DENADAI, B.S., FIGUERA, T.R., FAVARO, O.R.P. *et al*. Effect of the aerobic capacity on the validity of the anaerobic threshold for determination of the maximal lactate steady state in cycling. *Braz J Med Biol Res*. [online]. Oct. 2004, vol.37, no.10 [cited 25 January 2006], p.1551-1556.
10
FORJAZ, C.L.M., MATSUDAIRA, Y., RODRIGUES, F.B. *et al*. Post-exercise changes in blood pressure, heart rate and rate pressure product at different exercise intensities in normotensive humans. *Braz J Med Biol Res*. [online]. Oct. 1998, vol.31, no.10 [cited 25 January 2006], p.1247-1255.
11
ANTAS, P.R.Z., SALES, J.S., PEREIRA, K.C. *et al*. Patterns of intracellular cytokines in CD4 and CD8 T cells from patients with mycobacterial infections. *Braz J Med Biol Res*. [online]. Aug. 2004, vol.37, no.8 [cited 25 January 2006], p.1119-1129.
12
ANGELUCCI, M.E.M., CESARIO, C., HIROI, R.H. *et al*. Effects of caffeine on learning and memory in rats tested in the Morris water maze. *Braz J Med Biol Res*. [online]. Oct. 2002, vol.35, no.10 [cited 25 January 2006], p.1201-1208.
13
SUFFREDINI, I.B., SADER, H.S., GONCALVES, A.G. *et al*. Screening of antibacterial extracts from plants native to the Brazilian Amazon Rain Forest and Atlantic Forest. *Braz J Med Biol Res*. [online]. Mar. 2004, vol.37, no.3 [cited 25 January 2006], p.379-384.
14
MIYAMOTO, S.T., LOMBARDI JUNIOR, I., BERG, K.O. *et al*. Brazilian version of the Berg balance scale. *Braz J Med Biol Res*. [online]. Sept. 2004, vol.37, no.9 [cited 25 January 2006], p.1411-1421.
15
SERRAO, F.V., FOERSTER, B., SPADA, S. *et al*. Functional changes of human quadriceps muscle injured by eccentric exercise. *Braz J Med Biol Res*. [online]. June 2003, vol.36, no.6 [cited 25 January 2006], p.781-786.
16
MENEZES, J.R.L., MARINS, M., ALVES, J.A.J. *et al*. Cell migration in the postnatal subventricular zone. *Braz J Med Biol Res*. [online]. Dec. 2002, vol.35, no.12 [cited 25 January 2006], p.1411-1421.
17
NADER, H.B., PINHAL, M.A.S., BAU, E.C. *et al*. Development of new heparin-like compounds and other antithrombotic drugs and their interaction with vascular endothelial cells. *Braz J Med Biol Res*. [online]. June 2001, vol.34, no.6 [cited 25 January 2006], p.699-709.
18
MIRANDA, M.S., CINTRA, R.G., BARROS, S.B.M. *et al*. Antioxidant activity of the microalga Spirulina 16éd16ma. *Braz J 16éd Biol Res*. [online]. Aug. 1998, vol.31, no.8 [cited 25 January 2006], p.1075-1079.
19
MOLNAR, ªM., ALVES, ªA., PEREIRA-DA-SILVA, L. *et al*. Evaluation by blue native polyacrylamide electrophoresis colorimetric staining of the effects of physical exercise on the

activities of mitochondrial complexes in rat muscle. *Braz J Med Biol Res*. [online]. July 2004, vol.37, no.7 [cited 25 January 2006], p.939-947.
20
ANDRADE, L. et al . Psychometric properties of the Portuguese version of the State-Trait Anxiety Inventory applied to college students: factor analysis and relation to the Beck Depression Inventory. Braz J Med Biol Res., Ribeirão Preto, v. 34, n. 3, 2001.

Stemm Cells

1
Bongso, A., Fong, C.Y., Ng, S.C., and Ratnam, S.S. (1994). Blastocyst transfer in human in vitro: fertilization; the use of embryo co-culture. Cell Biol. Int. 18, 1181–1189.
2
Shamblott, M.J., Axelman, J., Wang, S., Bugg, E.M., Littlefield, J.W., Donovan, P.J., Blumenthal, P.D., Huggins, G.R., and Gearhart, J.D. (1998). Derivation of pluripotent stem cells from cultured human primordial germ cells. Proc. Natl. Acad. Sci. U. S. A. *95,* 13726–13731.
3
Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., Becker, R.A., and Hearn, J.P. (1995). Isolation of a primate embryonic stem cell line. Proc. Natl. Acad. Sci. U. S. A. *92,* 7844–7848.
4
Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science. *282,* 1145–1147.
5
M. Richards, C.Y. Fong and W.K. Chan *et al.*, Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells, *Nat Biotechnol* 20 (2002), pp. 933–936.
6
M. Amit and J. Itskovitz-Eldor, Derivation and spontaneous differentiation of human embryonic stem cells, *J Anat* 200 (2002) (Pt 3), pp. 225–232
7
M. Richards, S. Tan and C.Y. Fong *et al.*, Comparative evaluation of various human feeders for prolonged undifferentiated growth of human embryonic stem cells, *Stem Cells* 21 (2003), pp. 546–556.
8
Irina Klimanskaya1,2, Young Chung1,2, Sandy Becker1, Shi-Jiang Lu1 and Robert Lanza1. Human embryonic stem cell lines derived from single blastomeres. *Nature* advance online publication 23 August 2006
9
R. Mann, Imprinting in the germ line, *Stem Cells* 19 (2001), pp. 287–294
10
W.E. Wright and J.W. Shay, Telomere dynamics in cancer progression and prevention: fundamental differences in human and mouse telomere biology, *Nat Med* 6 (2000), pp. 849–851
11
M. Richards, S.P. Tan and J.H. Tan *et al.*, The transcriptome profile of human embryonic stem cells as defined by SAGE, *Stem Cells* 22 (2004), pp. 51–64.
12
M.J. Abeyta, A.T. Clark and R.T. Rodriguez *et al.*, Unique gene expression signatures of independently-derived human embryonic stem cell lines, *Hum Mol Genet* 13 (2004), pp. 601–608.
13
Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U. S. A. *78,* 7634–7638.
14
Yoshimizu, N. Sugiyama and M. De Felice *et al.*, Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice, *Dev Growth Differ* 41 (1999), pp. 675–684.
15
M. Pesce, X. Wang, D.J. Wolgemuth and H. Scholer, Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation, *Mech Dev* 71 (1998), pp. 89–98.
16
M.J. Shamblott, J. Axelman and S. Wang *et al.*, Human embryonic germ cell derivatives express a broad range of developmentally distinct markers and proliferate extensively in vitro, *Proc Natl Acad Sci USA* 98 (2001), pp. 113–118.
17
De Coppi P, Bartsch G Jr, Siddiqui MM, Xu T, Santos CC, Perin L, Mostoslavsky G, Serre AC, Snyder EY, Yoo JJ, Furth ME, Soker S, Atala A. Isolation of amniotic stem cell lines with potential for therapy. Nat Biotechnol. 2007 Jan;25(1):100-6.
18
Dor, J. Brown, O.I. Martinez and D.A. Melton, Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation, *Nature* 429 (2004), pp. 41–46.
19
C.R. Bjornson, R.L. Rietze and B.A. Reynolds *et al.*, Turning brain into blood: a hematopoietic fate adopted by adult neural stem cells in vivo, *Science* 283 (1999), pp. 534–537
20
K.A. Jackson, T. Mi and M.A. Goodell, Hematopoietic potential of stem cells isolated from murine skeletal muscle, *Proc Natl Acad Sci* 96 (1999), pp. 14482–14486.

Telomerase Group 1

1
Blackburn, E. H. & Gall, J. G. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J. Mol. Biol. 1978, 120:33-53.
2
Szostak, J. W. & Blackburn, E. H. Cloning yeast telomeres on linear plasmid vectors. Cell 1982, 29:245-255.
3
Murray, A. W. & Szostak, J. W. Construction of artificial chromosomes in yeast. Nature 1983, 305:189-193.
4
Shampay, J., Szostak, J. W., Blackburn, E. H. DNA sequences of telomeres maintained in yeast. Nature 1984, 310:154-157.
5
Dunn, B. L., Szauter, P., Pardue, M. L., Szostak, J. W. Transfer of telomere-adjacent sequences to linear plasmids by recombination. Cell 1984, 39:191-201.
6
Greider, C. W., & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 1985, 43:405-413.
7
Greider, C. W. & Blackburn, E. H. The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell 1987, 51:887-898.
8
Greider, C. W., & Blackburn, E. H A telomeric sequence in the RNA of Tetrahumena telomerase required for telomere repeat synthesis. Nature 1989, 337:331-337.
9
Lundblad V. & Szostak, J. W. A mutant with a defect in telomere maintenance leads to senescence in yeast. Cell 1989, 57:633-643.
10
Yu, G. L, Bradley, J. D., Attardi, L.D. and Blackburn, E. H. In vivo alteration of telomere sequences and senescence caused by mutated Tetrahymena telomerase RNAs. Nature 1990, 344:126-132.
11
Allsopp, R. C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, C. W., Greider, C. W., Harley, C. B. Telomere length predicts the replicative capacity of human

fibroblasts. Proc. Natl. Acad. Sci. USA 1992, 89:10114-10118.
12
Prowse, K. R., Avilion, A. A., Greider, C. W. Identification of a nonprocessive telomerase activity from mouse cells. Proc. Natl. Acad. Sci. USA 1993, 90:1493-1497.
13
McEachern, M. J. & Blackburn, E. H. Runaway telomere elongation cause by telomerase RNA mutations. Nature 1995, 376:403-409.
14
Rudolph, K. L., Chang, S, Lee, H.W., Blasco, M., Gottlieb, G., Greider, C. W., and DePinho, R. A. Longevity, stress response, and cancer in aging telomerase deficient mice. Cell 1999, 96:701-716
15
Kim, M. M., Rivera, M. A., Botchkina, I. L, Shalaby, R., Thor, A. D., Blackburn, E. H. A low threshold level of expression of mutant-template telomerase RNA is sufficient to inhibit tumor cell growth. Proc. Natl. Acad. Sci. USA 2001, 98:7982-7987

Telomerase Group 2

1
Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin,G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. Extension of life-span by introduction of telomerase into normal human cells. Science 1998, 279: 349–352.
2
Mitchell, J.R., Wood, E. & Collins, K. A telomerase component
is defective in the human disease dyskeratosis congenita. Nature 1999, 402: 551–555.
3
Chen, J.-L., Blasco, M.A., and Greider, C.W. Secondary structure of vertebrate telomerase RNA. Cell 2000, 100: 503–514.
4
Tzfati, Y., Fulton, T.B., Roy, J., and Blackburn, E.H. Template
boundary in a yeast telomerase specified by RNA structure. Science 2000, 288:863–867.

5
Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M.,Mason, P.J., and Dokal, I. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature 2001, 413:432–435.
6
Hermann, M. T., Strong, M. A. Hao, L. Y., Greider, C. W. (2001). The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. Cell, 107(1), 67-77.
7
Chen, J.-L., Opperman, K.K., and Greider, C.W. A critical stem-loop structure in the CR4–CR5 domain of mammalian telomerase RNA. Nucleic Acids Res. 2002, 30:592–597.
8
Seto, A.G., Livengood, A.J., Tzfati, Y., Blackburn, E.H., and Cech, T.R. (2002). A bulged stem tethers Est1p to telomerase RNA in budding yeast. Genes Dev., 2800–2812.
9
Ly, H., Xu, L., Rivera, M.A., Parslow, T.G., and Blackburn, E.H. A role for a novel "trans-pseudoknot" RNA-RNA-interaction in the
functional dimerization of human telomerase. Genes Dev. 2003, 17: 1078–1083.
10
Loayza, D., and de Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. Nature 2003, 423:1013–1018.
11
Chen, J.-L., and Greider, C.W. Template boundary definition of hTERT. Genes Dev. 2003, 17:2747–2752.
12
Armanios, M. et al. Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. Proc. Natl. Acad. Sci. USA 2005, 102:15960–15964.
13
Hao, L.Y. et al. Short telomeres, even in the presence of telomerase, limit tissue renewal capacity. Cell 2005, 123: 1121–1131.
14
Armanios, M. Y, Chen, J. J., Cogan, J. D., Alder, J. K., Ingersoll, R. G., Markin, C., Lawson, W. E., Xie, M, Vulto, I, Phillips, J. A., 16
Lansdorp, P. M., Greider, C. W., Loyd, J. E. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med. 2007, 356(13):1317-26.

Annex 2- Examples of the Analysis Form

ANALYSIS FORM	
Reference: Blackburn, E.H. and Gall, J.G. (1978) A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J. Mol. Biol.* 120: 33-53.	
ABSTRACT	
The extrachromosomal genes coding for ribosomal RNA (rDNA) in the ciliated protozoan *Tetrahymena thermophila* were studied with respect to sequences occurring at their termini. The linear rDNA molecules had previously been shown to be palindromic (Karrer and Gall, 1976; Engberg *et al.*, 1976). Within the terminal rDNA fragment produced by restriction endonuclease digestion, a tandomly repeated hexanucleotide sequence 5′ (C-C-C-C-A-A)$_n$ 3′ was found, where *n* is between 20 and 70. This fragment was heterogeneous in length as judged by gel electrophoresis. The repeating sequence was preferentially synthesized when rDNA was used as the template by *Escherichia coli* DNA polymerase I. Initiation occurred at specific single-strand discontinuities, probably one-nucleotide gaps, found every few repeats on the C-C-C-C-A-A strand. At least one discontinuity is present on the G-G-G-G-T-T strand. Experiments with T4 DNA polymerase suggested that there are no free cohesive ends on the rDNA of the kind found in bacteriophage λ DNA. The orientation of the strands carrying the repeating hexanucleotide sequence was determined, and a model for the termini of the rDNA based on these findings is presented.	
METHOD	
Theoretical Abductive:	**Experimental: Deductive: Inductive Exploratory X** EE
PROBLEM (transcribe from text)	
The sequence at the molecular structure at the termini of the extrachromosomal, linear, palindromic rDNA are therefore of considerable interest. (p. 34)	
OBJECTIVE	
Therefore, determination of the structure of the termini of the linear rDNA molecules should be useful for defining the processes involved in the completation of the replication in vegetative growing cells and in understanding the mechanism of application" (p. 34)	
HYPOTHESIS (transcribe from text)	
.	
CONCLUSIONS	
Within the terminal rDNA fragment produced by restriction endonuclease digestion, a tandomly repeated hexanucleotide sequence 5′ (C-C-C-C-A-A)$_n$ 3′ was found, where *n* is between 20 and 70.	
Normalized Relation: (determination of) the structure of the termini of the linear rDNA molecules: a tandomly repeated hexanucleotide sequence 5′ (C-C-C-C-A-A)$_n$ 3′ was found	
Antecedent/Mapping: the structure of the termini of the linear rDNA molecules	
Type of Relation/Mapping: has characteristic	
Consequent/Mapping: a tandomly repeated hexanucleotide sequence 5′ (C-C-C-C-A-A)$_n$ 3′ (was found) **MeSH terms:** Animals, Base Sequence, DNA Polymerase I, Electrophoresis, Nucleic Acid Denaturation, RNA, Ribosomal/*analysis/biosynthesis/genetics, Tetrahymena/*genetics/metabolism	

ANALYSIS FORM

Reference: Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P.J., and Dokal, I. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature *413*, 432–435, 2001.

ABSTRACT

Dyskeratosis congenita is a progressive bone-marrow failure syndrome that is characterized by abnormal skin pigmentation, leukoplakia and nail dystrophy. X-linked, autosomal recessive and autosomal dominant inheritance have been found in different pedigrees. The X-linked form of the disease is due to mutations in the gene DKC1 in band 2, sub-band 8 of the long arm of the X chromosome (ref. 3). The affected protein, dyskerin, is a nucleolar protein that is found associated with the H/ACA class of small nucleolar RNAs and is involved in pseudo-uridylation of specific residues of ribosomal RNA. Dyskerin is also associated with telomerase RNA (hTR), which contains a H/ACA consensus sequence. Here we map the gene responsible for dyskeratosis congenita in a large pedigree with autosomal dominant inheritance. Affected members of this family have an 821-base-pair deletion on chromosome 3q that removes the 3' 74 bases of hTR. Mutations in hTR were found in two other families with autosomal dominant dyskeratosis congenita.

METHOD

Theoretical Abductive	Experimental Deductive: Inductive X Exploratory: **EI**

PROBLEM (transcribe from text)

OBJECTIVE

- Here we map the gene responsible for dyskeratosis congenita in a large pedigree with autosomal dominant inheritance.
- The relative importance of rRNA processing and telomere maintenance in the pathophysiology of dyskeratosis congenita may be clari®ed by the nature of the genetic loci causing the autosomal form(s) of the disease.

HYPOTHESIS (transcribe from text)

- Dyskerin is also associated with telomerase RNA (hTR)5, which contains a H/ACA consensus sequence6,7.
- Defects in rRNA synthesis and/or in telomere maintenance might affect stem-cell function14.
- Dyskeratosis congenita patients have markedly shorter telomeres than normal individuals and this is apparent from an early age15.

CONCLUSIONS

Our ®nding of mutations in the telomerase RNA component (hTR) in three separate autosomal dominant pedigrees suggests that dyskeratosis congenita is due to defective telomerase activity

Normalized Relation/Mapping: mutations in the telomerase RNA component (hTR) causes dysqueratosis congenital

Antecedent/mapping: mutations in the telomerase RNA component (hTR) / Telomerase/*genetics, RNA/*genetics

Type of Relation/Mapping: causes / causes UMLS SN T147

Consequent/Mapping: dysqueratosis congênita / **Dyskeratosis Congenita/*genetics**

MeSH terms: Cell Line,Chromosome Mapping, *Chromosomes, Human, Pair 3, DNA Mutational Analysis, **Dyskeratosis Congenita/*genetics**, Female, Genes, Dominant, Humans, Linkage (Genetics), Male, *Mutation, Pedigree, Point Mutation, RNA/*genetics, Telomerase/*genetics, Telomere

ANALYSIS FORM

Reference: Seto, A.G., Livengood, A.J., Tzfati, Y., Blackburn, E.H., and Cech, T.R. A bulged stem tethers Est1p to telomerase RNA in budding yeast. Genes Dev. *16*, 2800–2812, 2002.

ABSTRACT

It is well established that the template for telomeric DNA synthesis is provided by the RNA subunit of telomerase; however, the additional functions provided by most of the rest of the RNA (>1000 nucleotides in budding yeast) are largely unknown. By alignment of telomerase RNAs of Saccharomyces cerevisiae and six Kluyveromyces species followed by mutagenesis of the S. cerevisiae RNA, we found a conserved region that is essential for telomere maintenance. Phylogenetic analysis and computer folding revealed that this region is conserved not only in primary nucleotide sequence but also in secondary structure. A common bulged-stem structure was predicted in all seven yeast species. Mutational analysis showed the structure to be essential for telomerase function. Suppression of bulged-stem mutant phenotypes by overexpression of Est1p and loss of co-immunoprecipitation of the mutant RNAs with Est1p indicated that this bulged stem is necessary for association of Est1p, a telomerase regulatory subunit. Est1p in yeast extract bound specifically to a small RNA containing the bulged stem, suggesting a direct interaction. We propose that this RNA structure links the enzymatic core of telomerase with Est1p, thereby allowing Est1p to recruit or activate telomerase at the telomere.

METHOD

Theoretical Abductive:	**Experimental: Deductive: Inductive Exploratory X** EE

PROBLEM (transcribe from text)

It is well established that the template for telomeric DNA synthesis is provided by the RNA subunit of telomerase; however, the additional functions provided by most of the rest of the RNA (>1000 nucleotides in budding yeast) are largely unknown.

… indicated that this bulged stem is necessary for association of Est1p, a telomerase regulatory subunit. (Abstract)

Identification of a structural element conserved among budding yeast telomerase RNAs (p.2802, Results)

An essential bulged stem in budding yeast telomerase RNA (p.2808, Discussio)

OBJECTIVE

In the present study, we describe the identification of an RNA bulged stem that is necessary for association of the telomerase regulatory protein Est1p with *S. cerevisiae* telomerase (p. 2800).

HYPOTHESIS (transcribe from text)

CONCLUSIONS

- The predicted bulged stem is essential for telomere maintenance in vivo (p.2802)
- Overexpression of telomerase protein Est1p but not Est2p suppresses the bulged-stem mutant phenotype (p. 2804).
- An intact bulged stem is required fo co-immunoprecipitation of telomerase RNA with the telomerase subunit Est1p but not Est2p (p. 2804).
- Interaction with Est1p is dependent on an intact bulged stem and independent of the full-length RNA (p.2809, Discussion)
- Conservation of the bulged stem in higher eukaryotes (p.2810, Discussion)

Normalized Relation: RNA Telomerase has characteristic a Budget stem

Antecedent/Mapping: Telomerase RNA

Type of Relation/Mapping: has characteristic

Consequent/Mapping: budget stem
MeSH terms: Conserved Sequence, Mutation, RNA, Fungal/*genetics, Saccharomyces cerevisiae/*genetics, Saccharomyces cerevisiae Proteins/*genetics, Sequence Alignment, Sequence Analysis, Telomerase/*genetics

ANALYSIS FORM

Reference: Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, Lawson WE, Xie M, Vulto I, Phillips JA 3rd, Lansdorp PM, Greider CW, Loyd JE. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med., 356 (13):1317-26, 2007.

ABSTRACT

Background: Idiopathic pulmonary fibrosis is progressive and often fatal; causes of familial clustering of the disease are unknown. Germ-line mutations in the genes hTERT and hTR, encoding telomerase reverse transcriptase and telomerase RNA, respectively, cause autosomal dominant dyskeratosis congenita, a rare hereditary disorder associated with premature death from aplastic anemia and pulmonary fibrosis.
Methods: To test the hypothesis that familial idiopathic pulmonary fibrosis may be caused by short telomeres, we screened 73 probands from the Vanderbilt Familial Pulmonary Fibrosis Registry for mutations in hTERT and hTR.
Results: Six probands (8%) had heterozygous mutations in hTERT or hTR; mutant telomerase resulted in short telomeres. Asymptomatic subjects with mutant telomerase also had short telomeres, suggesting that they may be at risk for the disease. We did not identify any of the classic features of dyskeratosis congenita in five of the six families.
Conclusions: Mutations in the genes encoding telomerase components can appear as familial idiopathic pulmonary fibrosis. Our findings support the idea that pathways leading to telomere shortening are involved in the pathogenesis of this disease.

METHOD

Theoretical Abductive: | **Experimental: Deductive:** **Inductive X Exploratory -** EE

. PROBLEM (transcribe from text)

Idiopathic pulmonary fibrosis is progressive and often fatal; causes of familial clustering of the disease are unknown. (Abstract)
the genetic basis of familial forms of idiopathic pulmonary fibrosis is not understood. [4] (p. 4)

OBJECTIVE

HYPOTHESIS (transcribe from text)

. familial idiopathic pulmonary fibrosis may be caused by short telomeres (Abstract)
- This pattern implies that in these patients, it is not the telomerase mutation itself but the short telomeres that determine the severity of the disease. [14,24,26] (p. 4)
- we hypothesized that telomere shortening causes this disease and that mutations in telomerase may contribute to it. (p. 4)

CONCLUSIONS

-We have shown that mutant telomerase is associated with familial idiopathic pulmonary fibrosis, which suggests that the spectrum of disease caused by telomere shortening is more **extensive t**han previously appreciated and that a subgroup of families with pulmonary fibrosis falls on that spectrum. (p. 9)
- There is evidence that short telomeres, rather than telomerase mutations, correlate with disease in dyskeratosis congenita. (p. 9)
- This observation is consistent with the variable penetrance associated with familial idiopathic pulmonary fibrosis and also suggests that the onset of disease may be age dependent. (p. 7)

Normalized Relation: shot telomeres may cause familial idiopathic pulmonary fibrosis

Antecedent/Mapping: short telomeres

Type of Relation/Mapping: may cause / causes UMLS SN T147

Consequent/Mapping: familial idiopathic pulmonary fibrosis **/ Pulmonary Fibrosis/*genetics/radiography**
MeSH terms: Female, Genes, Dominant, Heterozygote, Humans, Male, *Mutation, Mutation,Missense, Pedigree, **Pulmonary Fibrosis/*genetics/radiography**, RNA/*genetics, Reverse Transcriptase Polymerase Chain Reaction, Telomerase/*genetics/metabolism, Telomere/enzymology/genetics/*pathology