

This article was written for Semantic Universe website

## Agent Based Semantic Web

George Rzevski and Petr Skobelev

### Introduction

The Internet is a vast digital *network of computers* spreading around the globe. In 2008 the number of people connected to the Internet via servers, desktops and various mobile devices reached 1,463,632,361, which represents 21.9% of population of the world. One could argue that a quarter of all people who populate the Earth live in a *global village* – they can rapidly communicate with each other, exchange gossip, show photos, trade, provide services and ask each other for help.

The real breakthrough in establishing a useful global network came with the invention of the World Wide Web by Tim Berners-Lee in 1989. The Web is a *network of documents* in a standard format, stored on interconnected computers. In 2008 it was established that the indexable Web contains at least 63 billions web pages and Google announced that their search engine had discovered one trillion unique URLs. The significance of the Web is that it enables documents to be linked directly irrespective of their location.

The third stage in moving towards a true global village is well under way. The idea is to build a *network of content* stored on the web – the *Semantic Web* – making it possible for machines to understand meaning of data and to satisfy requests from people and machines to use the web content.

*Semantics* is the study of meaning in communication. The word derives from Greek *semantikos* "significant", from *semaino* "to signify, to indicate" and from *sema* "sign, mark, token". In linguistics it is the study of interpretation of signs as used by agents or communities within particular circumstances and contexts. It has related meanings in several other fields.

Tim Berners-Lee originally expressed his vision of the semantic web as follows:

"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The '*intelligent agents*' people have touted for ages will finally materialize" (Tim Berners-Lee, 1999).

Emergent Intelligence Technology Corporation has developed a version of intelligent agents mentioned by Tim Berners-Lee, agents capable of *semantic analysis* of web document content. The agent-based method for defining semantics enables computers to understand contents of documents written in a natural language such as English.

Possible applications of semantic analysis are numerous and include:

- Written communication between people and computers
- Written communication among computers
- Software translators
- Text referencing engines
- Semantic search engines
- Auto-abstracting engines
- Annotation and classification systems
- Semantic document-flow management systems

Despite a considerable research effort in areas such as computer linguistics, artificial intelligence and neural networks the problem of text understanding by computers has not been effectively solved. The reason may well be that the currently proposed solutions to this problem are strictly centralised, sequential and static. In contrast, the method described in this article is based on the concept of autonomous software agents dynamically co-operating, competing or arguing with each other and, through a process of pro-active negotiations, refining tentative semantic solutions until an agreed semantics of the text is established.

The main idea of the new approach is that a software agent is assigned to each word of the text under consideration. Agents have access to a comprehensive repository of knowledge about possible meanings of words in the text and engage into negotiation with each other until a consensus is reached on meanings of each word and each sentence. In some cases the method may discover several contradictory meanings of a sentence. The conflict is then resolved by an agent-triggered consultation with the user and consequent updating of the repository of knowledge. To simplify the process of extracting meanings, the method performs an initial morphological and syntactic analysis of the text.

## Definitions

Key concepts of the proposed method are as follows.

*An Agent* is a software object capable of contributing to the accomplishment of a task by

- Accessing domain knowledge
- Reasoning about it's task
- Composing meaningful messages
- Sending them to other agents or humans
- Interpreting received messages
- Making decisions based on domain knowledge and collected information
- Acting upon decisions in a meaningful manner

*A Multi-Agent System* is a software system consisting of agents competing or co-operating with each other with a view to accomplishing system tasks. The main principle of achieving goals within such system is a negotiation among agents, aimed at finding a balance between many different interests of individual agents.

*Ontology* is a conceptual description of a domain of the Universe under consideration. Concepts are organised in terms of objects, processes, attributes and relations. Values defining instances of concepts are stored in associated databases. Concepts and values together form the domain knowledge.

*A Syntactic Descriptor* is a network of words linked by syntactic relations representing a grammatically correct sentence.

*A Semantic Descriptor* is a network of grammatically and semantically compatible words, which represents a computer readable interpretation of the meaning of a text. If semantic ontology describes all possible meanings of words in a domain, a semantic descriptor describes the meaning of a particular text.

*Self-organisation* is the capability of a system to autonomously, ie, without human intervention, modify existing and/or establish new relationships among its components with a view to increasing a given value or recovering from a disturbance, such as, an unexpected addition or subtraction of a component. In the context of text understanding any autonomous change of a link between two agents representing different meanings of words is considered as a step in the process of self-organisation.

*Evolution* is the capability of a system to autonomously modify its components and/or links in response, or in anticipation of changes in its environment. In the context of text understanding any autonomous update of Ontology based on the newly acquired information is considered as a step in the process of evolution.

## **The Agent-Based Method for Semantic Analysis**

The method consists of the following four steps:

1. Morphological analysis
2. Syntactic analysis
3. Semantic analysis
4. Pragmatics

The text is divided into sentences. Sentences are fed into the meaning extraction process one by one.

### **Morphological Analysis**

1. An agent is assigned to each word in the sentence
2. Word Agents access Ontology and acquire relevant knowledge on morphology
3. Word Agents execute morphological analysis of the sentence and establish characteristics of each word, such as gender, number, case, time, etc.
4. If morphological analysis results in polysemy, ie, a situation in which some words could play several roles in a sentence (a noun or adjective or verb), several agents are assigned to the same word each representing one of its possible roles

## Syntactical Analysis

5. Word Agents access Ontology and acquire relevant knowledge on syntax
6. Word Agents execute syntactical analysis where they aim at identifying the syntactical structure of the sentence. For example, a Subject searches for a Predicate of the same gender and number, and a Predicate looks for a suitable Subject and Objects. Conflicts are resolved through a process of negotiation. A grammatically correct sentence is represented by means of a Syntactic Descriptor
7. If results of the syntactical analysis are ambiguous, ie, several variants of the syntactic structure of the sentence under consideration are feasible, each feasible variant is represented by a different Syntactic Descriptor

## Semantic Analysis

8. Word Agents access Ontology and acquire relevant knowledge on semantics
9. Each grammatically correct version of the sentence under consideration is subjected to semantic analysis. This analysis is aimed at establishing the semantic compatibility of words in each grammatically correct sentence. Word Agents learn from Ontology possible meanings of words that they represent and by consulting each other attempt to eliminate inappropriate alternatives
10. Once agents agree on a grammatically and semantically correct sentence, they create a Semantic Descriptor of the sentence, which is a network of concepts and values contained in the sentence
11. If a solution that satisfies all agents cannot be found, agents compose a message to the user explaining the difficulties and suggesting how the issues could be resolved
12. Each new grammatically and semantically correct sentence generated by the steps 1 - 11 is checked for semantic compatibility with Semantic Descriptors of preceding sentences. In the process agents may decide to modify previously agreed semantic interpretations of words or sentences (self-organisation)
13. When all sentences are processed, the final Semantic Descriptor of the whole document is constructed thus providing a computer readable semantic interpretation of the text

## Pragmatics

14. Word Agents access Ontology and acquire relevant knowledge on pragmatics, which is closely related to the application at hand
15. At this stage agents consider their application-oriented tasks and decide if they need to execute any additional processes. For example, if the application is a

Person – Computer Dialog, agents may decide that they need to ask the user to supply some additional information; if the application is a Search Engine, agents will compare the Semantic Descriptor of the search request with Semantic Descriptors of available search results. If the application is a Classifier, agents will compare Semantic Descriptors of different documents and form groups of documents with semantic proximity.

Let us recapitulate main features of the proposed method.

- Decision making rules are specified in ontology, which incorporates general knowledge on text understanding, language-oriented rules and specific knowledge on the problem domain
- Every word in the text under consideration is given the opportunity to autonomously and pro-actively search for its own meaning using knowledge available in ontology
- Tentative decisions are reached through a process of consultation and negotiation among all Word Agents
- The final decision on the meaning of every word is reached through a consensus among all Word Agents
- Semantic Descriptors are produced for individual sentences and for the whole text
- The extraction of meanings follows an autonomous trial-and-error pattern (selforganisation)
- The process of meaning extraction can be regulated by modifying ontology

### **An Example of Semantic Analysis**

The proposed method has been applied to the problem of searching for relevant abstracts.

Fig. 1 shows a published abstract of a scientific paper, which needs to be converted into a computer readable format using the method described in this article.

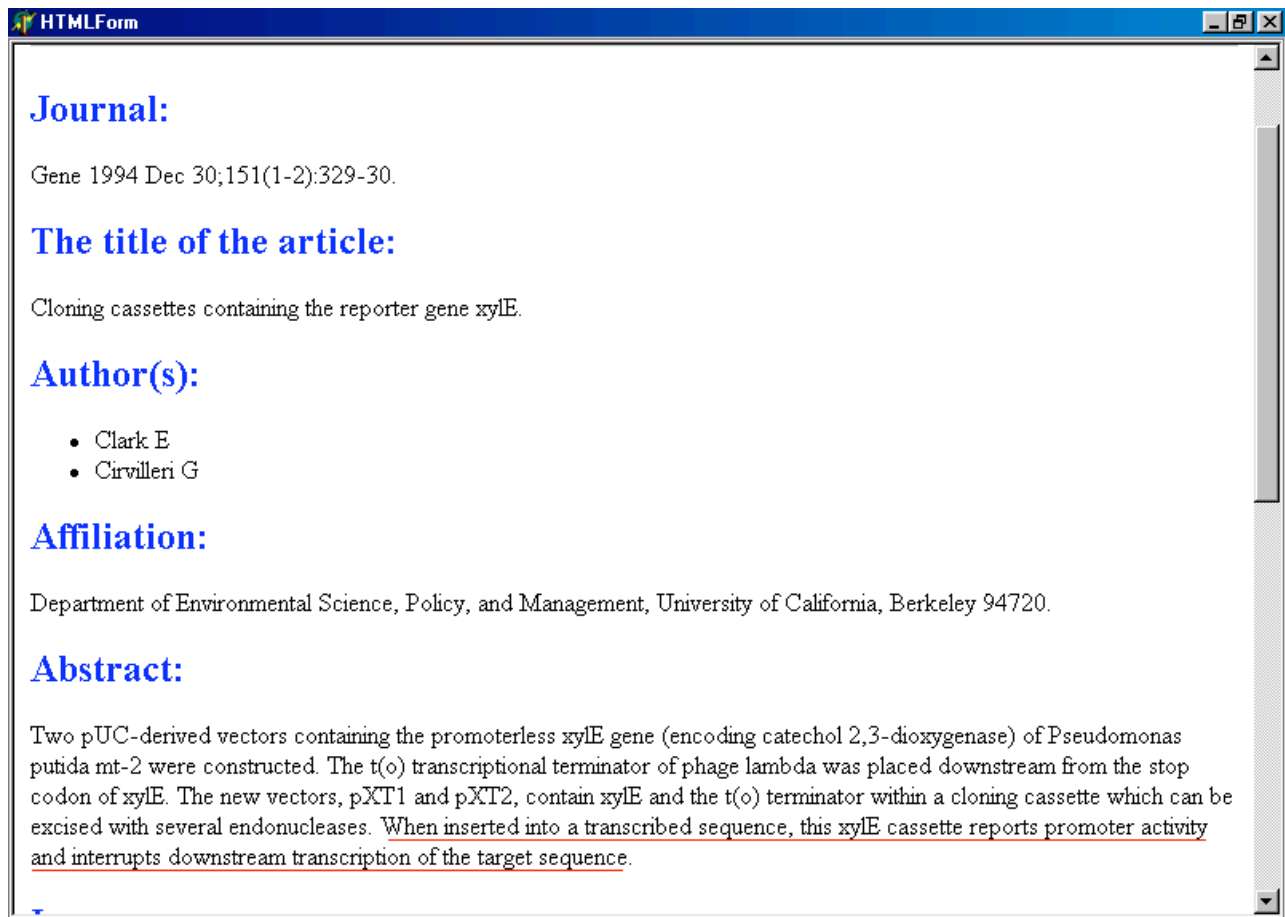


Fig.1 A text of a selected article

The semantic descriptor of the title of the abstract is shown in Fig. 2a Note that the sentence has been completely understood by the system - the relations between the gene and locus and gene properties have been determined. Note that their meaning are shown at the bottom of the screen (in biology a locus by definition is a specific site of the particular gene or chromosome; according to domain ontology "cloning cassette" is a synonym of the semantic concept "locus").

Fig. 2b shows how a tentative semantic descriptor of the whole text is modified during semantic analysis in a stepwise manner. Blue links indicate connections that were added to the semantic descriptor during the analysis of the last sentence of the text (the underlined sentence from Fig.1). As a result of the analysis of the last sentence the system discovered some new concepts and new relations between the existing nodes of the descriptor, including, a new relation «Have» between the gene and the locus; furthermore, the gene has obtained a new Insert relation and the relation «Have» has been established between the locus and the new node, operon (in biology operon by definition is a controllable unit of transcription consisting of a number of structural genes transcribed together; it contains at least two distinct regions: the operator and the promoter; therefore, according to ontology and the text of the abstract, the semantic descriptor includes the concept "operon").

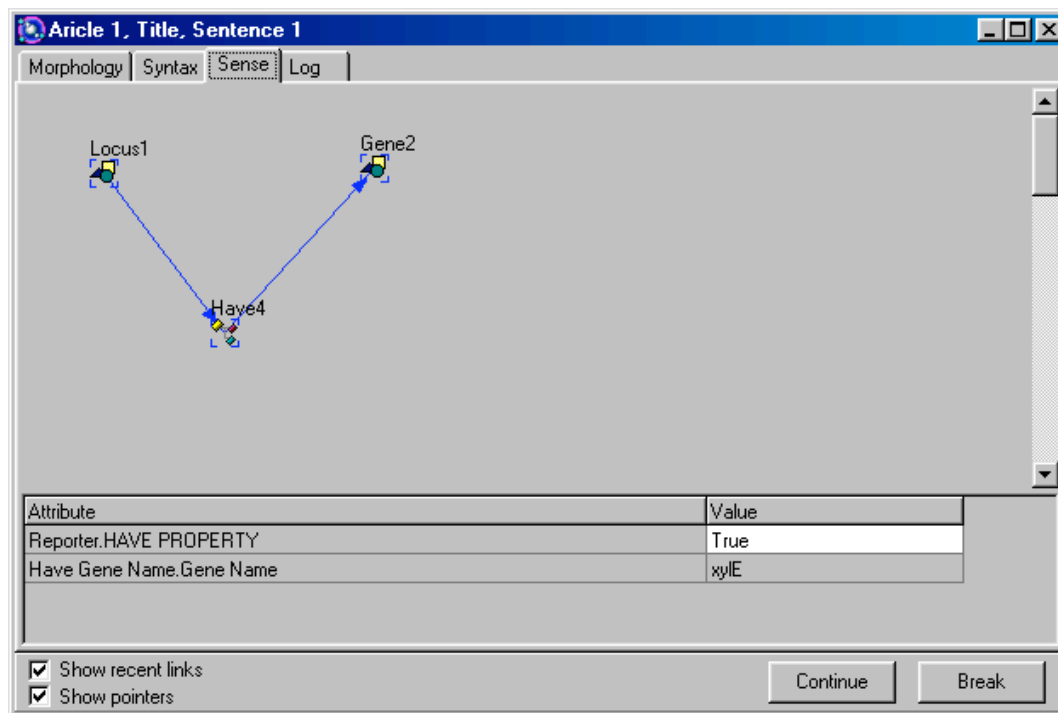


Fig.2a Semantic descriptors for the title

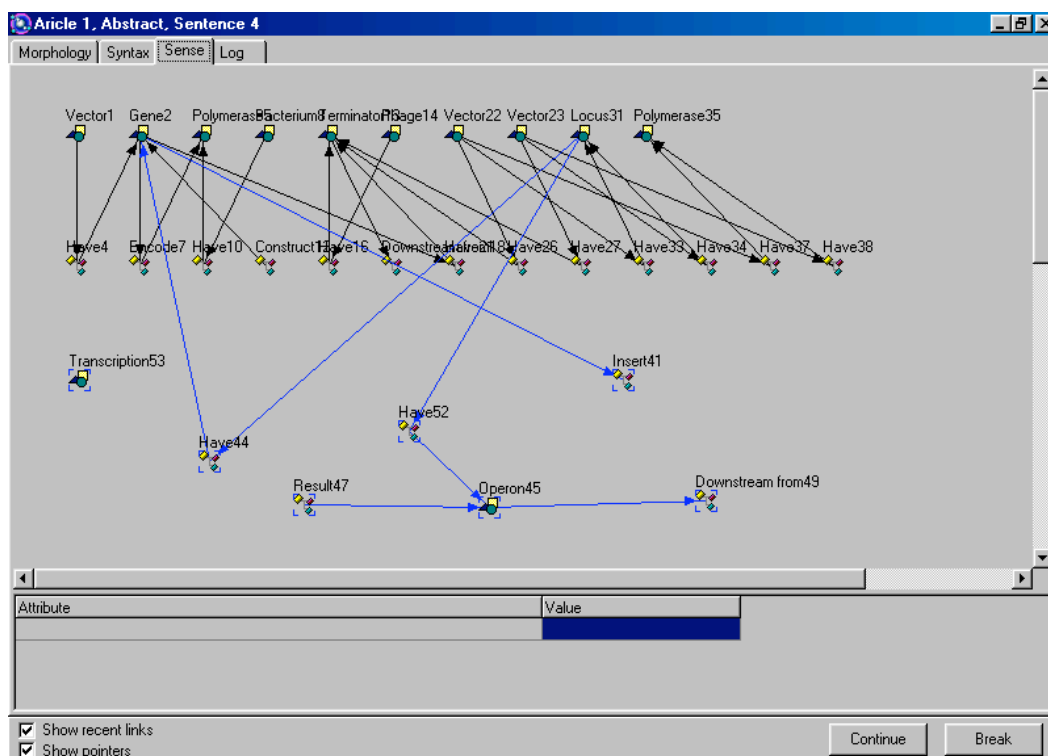


Fig.2b Semantic descriptors for the text

The final semantic descriptor of the whole abstract is shown in Fig.3

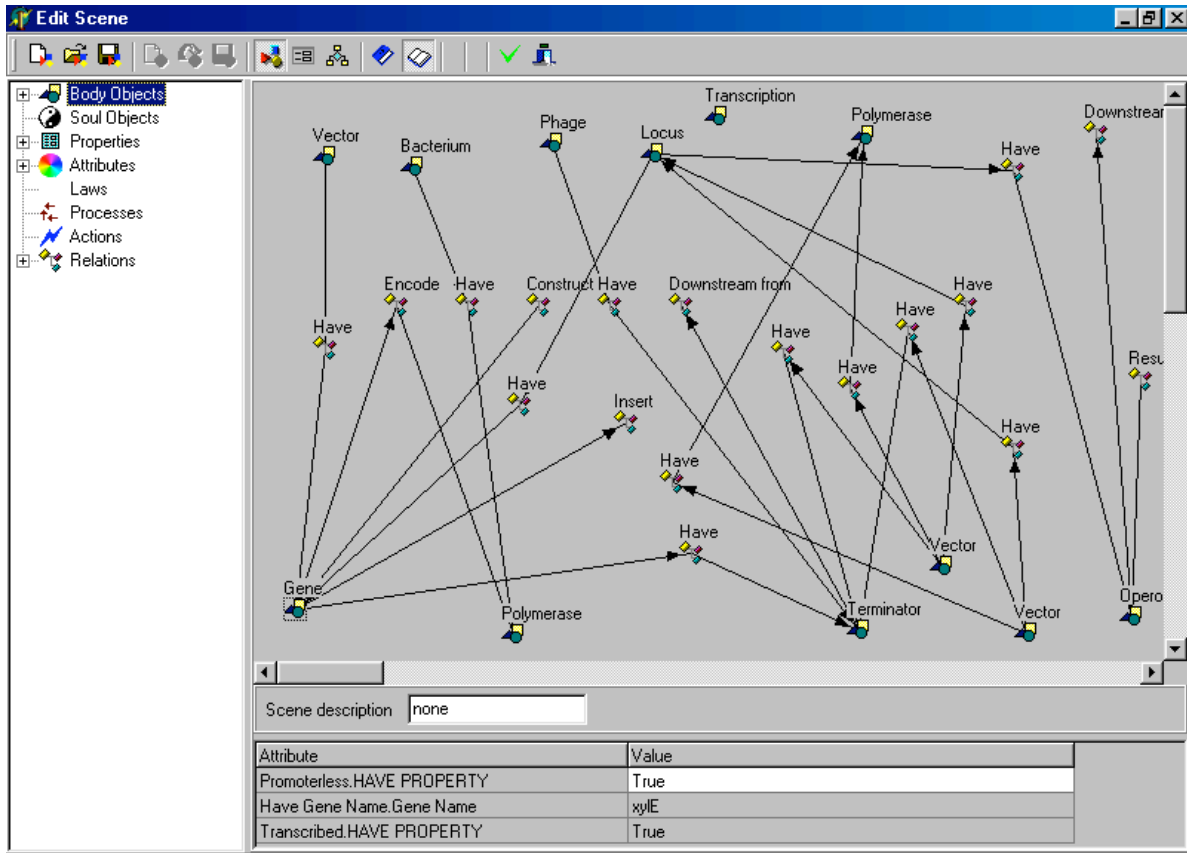


Fig.3 - Semantic descriptor of the abstract from Fig. 1

In addition to creating semantic descriptors for each abstract it is necessary to formulate a semantic descriptor of the enquiry. Fig.4 shows a semantic descriptor of a request to search for abstracts in which an organism is connected with a sequence through the relation Have.

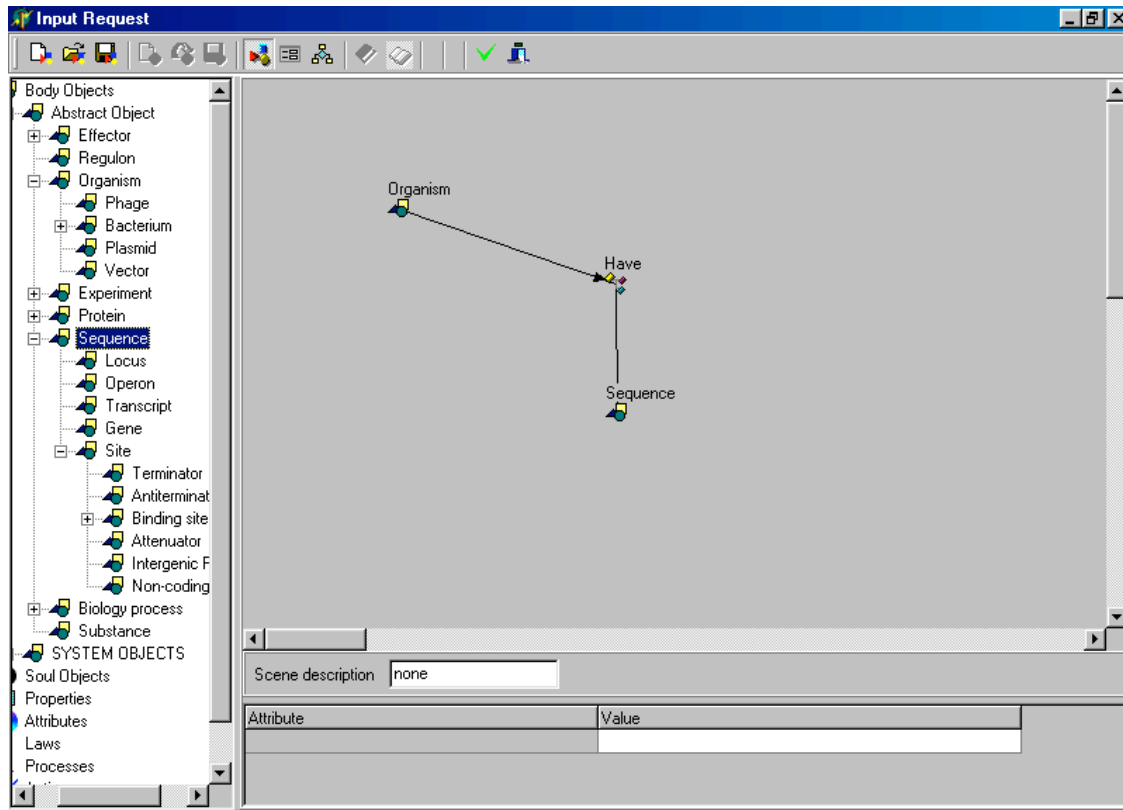


Fig.4 A request to search for abstracts with a particular content

In Fig.5 the best matching abstract is marked in blue; yellow denotes all abstracts, which match the request. All the comparisons are made based on rules specified in ontology. A change in ontology may change the ranking of semantic descriptors.

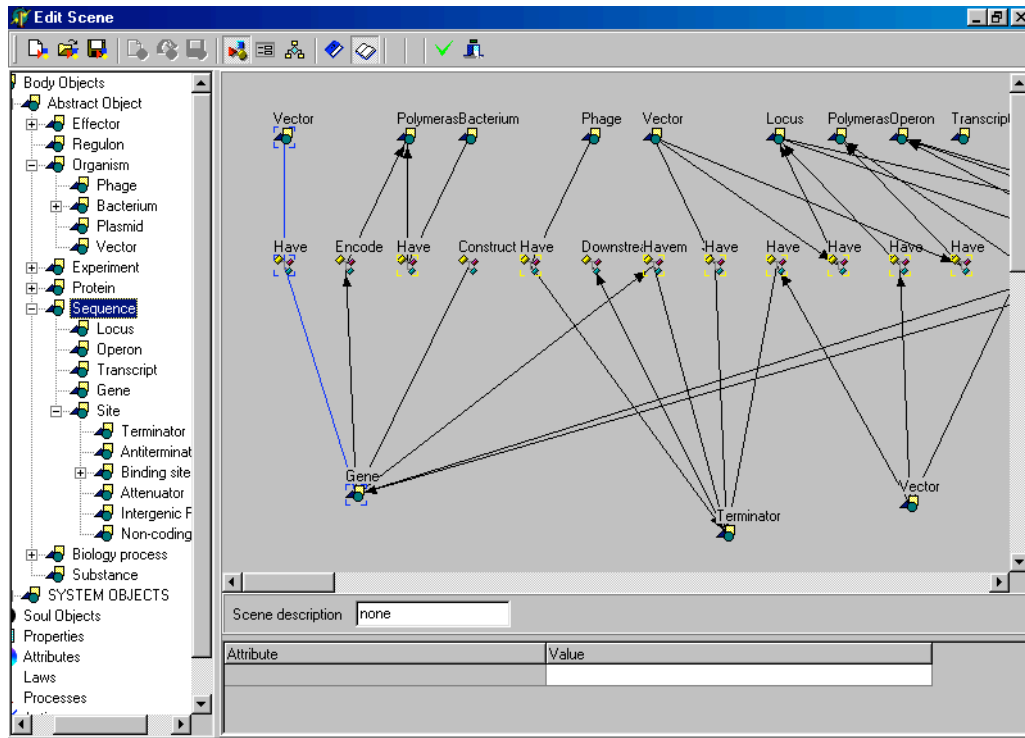


Fig.5 Comparison of semantic descriptors of analysed abstracts

## Acknowledgement

It is our pleasant duty to acknowledge the contributions to the development of this new method for semantic analysis by Dr Igor Minakov who has solved the problem of semantic matching of abstracts to queries described in this article.

## Conclusion

Autonomous agents offer an effective way of allocating meanings to words primarily because the search algorithms are replaced by broadcasting of messages. Distributed decision making by agents assigned to words enables fast discovery of context within which text can be understood.