УДК 613.648

# СВЯЗАННЫЕ ПОТРЕБНОСТИ

**© 2010 г.** ***А. М. Абрамович***

***Ведущая государственная лаборатория по разработке программного обеспечения, Уханьский университет, Китай***

*Современные веб-технологии эволюционируют в сторону новой виртуальной реальности, где вместо обычных сайтов и страниц пользователю интернета будет предоставляться структурированная аудио-, видео- или текстовая информация. Эта информация будет генерироваться автоматически в качестве ответа на представленный список потребностей пользователя. Создавшаяся ситуация требует возникновения различных целевых веб-поисковиков, которые предоставят предметно-ориентированные интерфейсы для получения и обработки запросов пользователей.*

*В настоящей статье рассмотрен новый подход, направленный на извлечение, представление и просмотр знаний о потребностях человека во Всемирной Сети.*

Ключевые слова: *информационные потребности и запросы; связанные данные; связанные потребности; онтология; семантика; представление знаний; ризонинг; вопросно-ответная система.*

*Modern Web technologies lead to a new virtual reality, where instead of the usual sites and pages Internet user will be provided with structured audio, video or textual information. This information will be generated as a response to the specification of user's needs. The situation requires various Target Web search engines to appear that provide domain-specific interfaces for the obtaining and processing of users specifications. Target Web search engines shall process commonsense knowledge, which expresses human needs to be mined from different sources by different ways.*

*This paper offers an overview of the new approach aimed at extraction, representation and browsing knowledge about human needs in the Web.*

Key words: *Linked Data; Linked Needs; ontology; semantics; knowledge representation; reasoning; query-answering system.*

## 1. Introduction

The original idea of the Semantic Web is based upon three foundations, namely: RDF, XML and ontological pages [1]. After 10 years of research, the SW was rebranded into Web of Data. Constructive material of Web of Data is Linked Data. "Linked Data is simply about using the Web to create typed links between data from different sources" [2].

The Linked Data tenets [1; 2] are formulated by Tim Berners-Lee as a following:

1. Use URIs (Uniform Resource Identifiers) as names for "things".

2. Use HTTP URIs so that people can look up those names.

3. When someone looks up a URI, provide useful information.

4. Include links to other URIs, so that they can discover more things.

Linked Data browsers will provide user with configurations of RDF triples that connect different data sources into a unique segment of information adequate to the user's specification.

By Web of Data notation Web publisher provides a set of RDF triples that represents an adequate configuration of its domain.

At present time a total number of Web located RDF triples is almost 4 billion.

How is possible to specify the need, by means of which the engine will select and give the relevant data? To use existing relational databases one needs to know what information they contain, and how. Knowing this, users compose their queries.

But Web of Data has not a fixed structure.

How to specify a query to Web of Data?

Leading researchers in the field of Semantic Web and Linked Data (Tom Heath, Bill Baxton, etc.) consider that "diversity of 'Web browsers' tomorrow will match the diversity of 'ink browsers' (a.k.a. paper) today — in terms of diversity of form, function, location, and importance" [5]. Such point of view exists due to a necessity to configure the optimal subset of all RDF triples related to things that compose user needs' specifications. This implies that both user interface and the browsing must be user need oriented.

Existing Web of Data browsers, for example, Tabulator [6; 7] and DBpedia Mobile [8; 10] provide a possibility for users to set a field of their interests in a personal table (Tabulator) or to use a fix set of data fields (DBpedia Mobile).

Possibilities of DBpedia as a whole are restricted by DBpedia Knowledge Extraction Framework [9] that includes a fix set of search parameters.

Unfortunately, such approaches guarantee getting only discrete, referential and selective information, whereas an end user needs to receive comprehensive data related to his problem.

The solution to this problem could be a unified semantic framework that grounds both the structuring of Linked Data objective segments and the capturing of user need's specification.

This article documents the recent progress in the effort to build a system of commonsense knowledge design as well as commonsense reasoning representation. It builds upon two earlier publications about the project [3; 4]. This project deals with a realization of Need Language (NL language).

NL language provides an information extraction framework that converts Web of Data content into a rich multi-domain knowledge base as an independent Web resource as well as a part of Linked Data named Linked Needs (LN).

The rest of article is structured as follows: We give an overview of the NL program architecture and knowledge extraction techniques in Section 2. The LN system is described in Section 3. Section 4 concludes and outlines future work.

## 2. The Linked Needs Knowledge Extraction Framework.

Any Web publisher has in mind a satisfaction of certain needs of Internet users. Any Internet user looks for in Internet the satisfaction of his needs. Internet user's problem is a discovery of relevant data in the Web that satisfies or helps to satisfy his needs. Web publisher's problem is giving the information in a form attractive to Internet users, whose needs he is ready to satisfy.

Information oversaturation of the Web has reached such proportions that Internet users need more and more time to find the necessary data, and Web publishers need more time and money to attract the attention of Internet users.

This dramatic situation could be solved if the Internet users will be able to clearly specify their needs (or, at least, their current situations), and Web browser will look for relevant information in a field of Linked Needs' representations (Linked Needs).

It is clear that both interaction with Internet users and Linked Needs must be ground on the same semantic framework.

NL language provides such framework.

### *2.1. NL language*

The NL language is a knowledge representation language for the description of customers' needs. It has the following advantages:

To give an opportunity for professionals in various subject areas to represent their experience in a form that is equally accessible both to the automatic semantic search system and to customers.

To give an opportunity both for professionals and laymen to use their customized slang to communicate with the computer; in particular, this concerns the description of the current situation and its forerunner.

To ground a representation of a Need Satisfaction Domain for any private person, for any professional or for the society as a whole.

To provide an opportunity of building a Global Knowledge Platform that supports different social activities.

To represent data as a program, and to represent a program as data with the purpose of solving problems using the equation

**Decision=Recognition + Synthesis + Control.**

The NL language is rather designed as an intermediate semantic markup language for capturing and accumulating knowledge and for maintaining man-machine communications, than as a means to directly control the computer, as done in a programming language.

Syntax of NL is used for guiding an intelligent interface with the following purposes:

• discovering the semantics of the current situation of the customer,

• determination of the current needs,

• separation of the real needs from the imaginary ones,

• offering different options for adequate behavior, and

• simulating different behaviors in the light of possible consequences.

*2.2. Linked Needs Semantic Framework*

LN Semantic Framework is based on the NL language notation [3; 4] and consists of components that define location (both Web and semantic) and semantic essence of the following types of resources:

• missing need,

• available resources, and

• constructive resources.

Semantic location is defined by means of an enumeration of two kinds of needs, namely, needs that antecede to the above-mentioned resources and needs that are logically derivative from them. Web location of resources is defined by URIs.

Missing need is defined by the name of the need, by synonyms of the need, by the type of the need and by list of sub-needs.

Available resources contain sets of resources that are necessary for the missing need's satisfaction include time resources, place resources, a budget of the missing need's satisfaction, necessary equipment, necessary materials, skills and so on. Constructive resource specifies the way of the missing need's satisfaction.

According to NL language notation the first semantic sharing and ontology of domain segment's representation is generic. This implies that any posterior semantic sharing and ontology of the same domain segment's representations must be harmonized with the generic one.

*2.3. Linked Needs' extraction framework*

Linked Needs' representations (Linked Needs) can be mined by the following ways:

1. Writing NL program;

2. Interaction with Web publisher;

3. Interaction with domain expert;

4. Semi-automatic extraction of need related information from Web resources.

Every of the above-mentioned Linked Needs' mining approaches results in NL programs and turns them into RDF triples.

Correspondently, there are four kinds of Linked Needs extractors (Figure 1):

1. NL editor;

2. Web publisher query-answering interface (Web publisher GAI);

3. Domain expert query-answering interface (Domain expert GAI);

4. Semi-automatic objective search engine (SAO engine).

**NL editor is simulation software** that allows assembling NL program using a panel of NL's constructs provided by NL templates and by query-answering interactions with appropriate segments of Linked Needs. In the scope of interaction with Linked Needs the system engine verifies of the published experience and retrieves its ontology and semantics online.

**Web publisher's GAI** results NL program provided by an ordered textual, graphical, audio and video content of his Web resource.

**Domain expert's GAI** results NL program.

**SAO** engine mines content for NL programs from Web resources, which are detected by Web search engine in response to key words.

**Internet user's GAI** results NL program.

In the scope of interaction via above-mentioned GAIs the query-answering reasoning parses the interlocutor's lexicon and harmonizes

it with generic ontology.

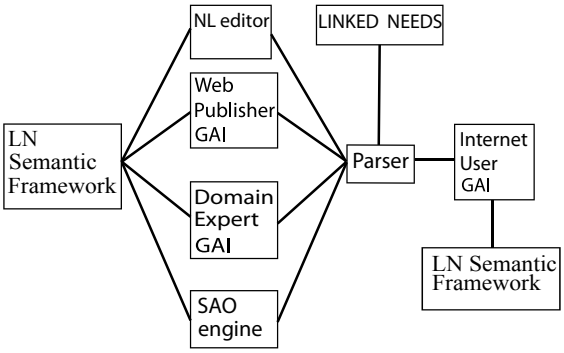


**Fig. 1**. Architecture of the Extraction Framework

*2.4. Linked Needs as a part of Linked Data*

Any NL program represents a semantically ordered set of information about three kinds of resources, namely, the missing resource, the available resource, and the constructive resource. This structure grounds a transformation of NL program into RDF triple, where a missing resource plays a role of the subject, an available resource plays a role of the object, and a the constructive resource plays role of predicate.

Any NL program describes a certain way of the given missing need's satisfaction. Every missing need may be satisfied in different ways. All available ways of the missing need satisfaction constitute the Need Satisfaction Domain [3; 4].

Linked Needs joins all Need Satisfaction Domains represented by means of RDF triples into one directed graph.

Linked Needs is a sub-topic of the Linked Data. The term Linked Needs is used here to describe a method of exposing, sharing, and connecting needs via deferencable URIs on the Web.

**3. SAO engine.**

The focus of this article is SAO engine since today's human-readable Internet contains a tremendous amount of useful information, which it's necessary to transform into Linked Needs. SAO engines a man-machine activity (Figure 2). A man is responsible for gathering of a set of specific typical speech cliches that together contain data covering the given Need Satisfaction Domain. A machine is responsible for the search in the texts of Web documents typical speech cliches as well as for mining of the required data and structuring it by means of the LN Semantic Framework. A man (Need Satisfaction Domain administrator) is responsible also for the editing of resulting NL programs.

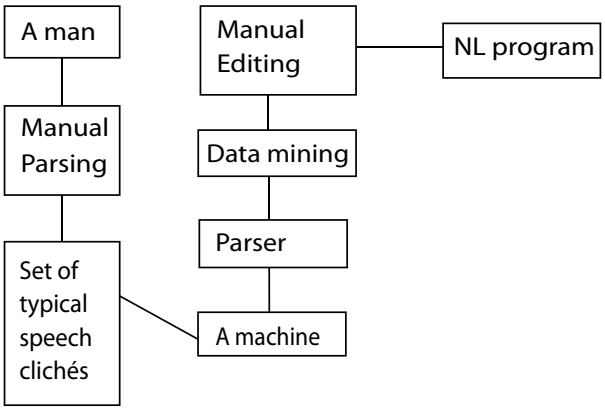


**Fig. 2.** SAO engine

Basic idea of SAO engine is as follows:

1. Each text contains constructive information that defines the need expressed by it or specifies a missing resource as a need.

2. This information is not a translation of the text, but it is a translation of the text to the language of human needs.

3. LN Semantic Framework provides "a semantic mask", which is used for extraction of the description of professional or other requirements of the person.

4. The "semantic mask" is domains' invariant, but every domain has its own particular forms (its typical speech cliches) for needs' expressing.

5. A human manually finds as many as possible typical speech cliches that cover the needs' expression of the given domain.

6. The system engine processes the given text, finds typical speech cliches and returns their contents.

7. The captured knowledge, next, is automatically represented in a machine-readable format by means of semantic marking according to NL language syntax.

8. After that the system engine checks the semantic completeness and clearness of received NL program (or its part). If, for example, the captured information contains terms unknown to system, it automatically accompanies them by questions, which the system administrator is obliged to answer, correcting thereby received NL text.

*3.1. Manual parsing*

The Linked Needs administrators analyze as many Web documents and form sets of typical speech cliches for every domain segment of Linked Needs as possible.

Let's consider medicine (Figure 3) and cooking (Figure 4) domains as examples, since their Web documents, as a rule, are structured.

Manual parsing detects the following and others typical speech cliches of medicine related documents:

Symptoms of NameOfDiseas: …

Symptoms of NameOfDiseas are …

The most characteristic symptoms of NameOfDiseas are …

The following are symptoms that may occur in specific types of NameOfDiseas: …

NameOfDiseas treatment requires the following medicines: ...

NameOfDiseas treatment requires the following equipment: ...

We treat NameOfDiseas by means of the following: ...

NameOfDiseas treatment consists of the following stages: ...

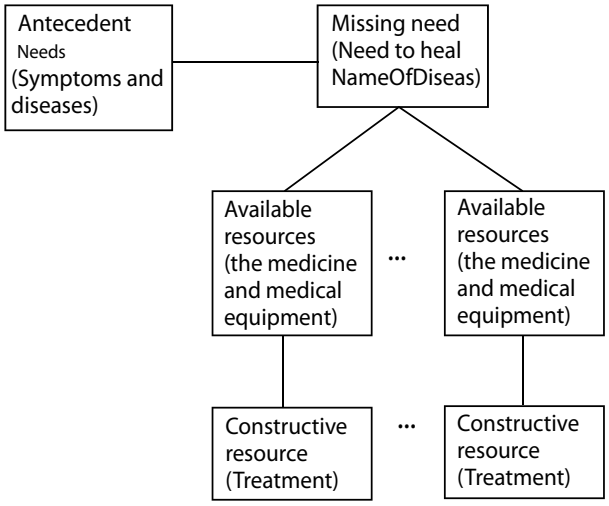


**Fig. 3.** Medicine Web data content

Manual parsing results the following and others typical speech cliches of cooking related recipes:

NameOfDush recipe:

NameOfDush ingredients are …

Ingredients: …

For NameOfDush use:

Cook time is…

Methods: …

Preparation: …

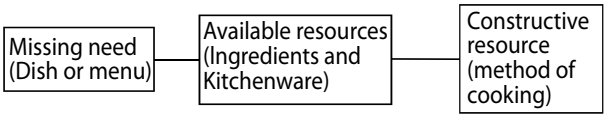


**Fig. 4.** Cooking Web data content

Cooking instructions: …

Recipe instructions: …

*3.2. Mining of ordered information from a news line*

Any human purposeful activity may be considered as a satisfaction of a certain need. A special subset of LN Semantic Framework, named Activity NL (ANL), grounds the human purposeful activity's description.

Figure 5 demonstrate the SAO engine process for mining of ordered information from a news line:

1. An editor of LN news line defines the basis social activities, information about which is interested for LN news line's readers.

2. The system administrator is responsible for both the generic representation of selected human activities and the formation of sets of the corresponding textual cliches that cover these representations.

3. The system engine parses the input text with the purpose of a discovering of ordered news.

4. Resulted information is edited by the newsmaker.

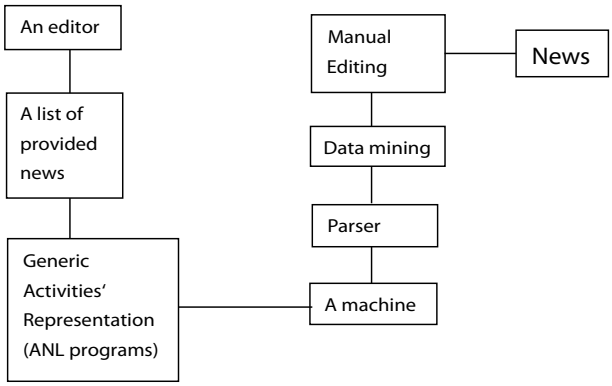


**Fig. 5.** Mining of ordered information from a news line

## 4. Conclusion

Linked Needs is Web resource that grounds a new virtual reality, namely, **Reality of needs' satisfaction**. Linked Needs contains collections of typical speech cliches, generic human activities' representation and Need Satisfaction

Domains both in RDF form and as NL programs. Since NL language [3; 4] is tailored according to the commonsense reasoning, Web applications that realize the commonsense reasoning will provide a satisfaction of Internet users' needs in form of tutorials, instructions, well-ordered information packages contained Web data in all allowable formats as well as in results of cloud computing.

## References

1. *Berners-Lee T.* Linked Data – Design Issues // World Wide Web Consortium [Электронный ресурс]. – 2006. – Электрон. дан. – Режим доступа: http://www.w3.org/DesignIssues/LinkedData.html, свободный. – Загл. с экрана.

2. *Bizer C., Cyganiak R., Heath T.* How to publish Linked Data on the Web // WWW4, the research application server of the Lehrstuhl für Wirtschaftsinformatik. [Электронный ресурс]. – 2007. – Электрон. дан. – Режим доступа: http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/, свободный. – Загл. с экрана.

3. *Abramovich A.*, Xu C. Z. et. al. The Need Language. – Collin2010, 2010.

4. *Abramovich A.*, Wang Q., Sheu P. C.-Y. Towards a Global Platform of Human Experiences. – SDPS2010, 2010.

5. *Heath T.* How Will We Interact with the Web of Data? // CS Digital Library. [Электронный ресурс]. – Электрон. дан. – Режим доступа: http://translated.by/you/how-will-we-interact-with-the-web-of-data/original/, свободный. – Загл. с экрана.

6. *Berners-Lee T.*, et. al. Tabulator Redux: Writing Into the Semantic Web (preprints). – 2007.

7. *Berners-Lee T.*, et. al. Tabulator: Exploring and Analyzing linked data on the Semantic Web // Proceedings of The 3rd International Semantic Web User Interaction Workshop (SWUI06). – Athens, Georgia. – 6 Nov 2006.

8. New Client Application // DBpedia Homepage. [Электронный ресурс]. – Электрон. дан. – Режим доступа: http://wiki.dbpedia.org/DBpediaMobile, свободный. – Загл. с экрана.

9. *Bizer C., Lehmann J., Kobilarov G.* et. al. DBpedia – A Crystallization Point for the Web of Data // Journal of Web Semantics: Science, Services and Agents on the World Wide Web, Issue 7, 2009.

10. *Becker C., Bizer C.* DBpedia Mobile: A Location-Enabled Linked Data Browser // 1st Workshop about Linked Data on the Web (LDOW2008). – Beijing, China. – April 2008.

*Поступила в редакцию* *18 апреля 2010 г.*

**Александр Марианович Абрамович –** доктор математических наук, исследователь Ведущей государственной лаборатории по разработке программного обеспечения Уханьского университета (Китай). Автор более 50 научных публикаций. Руководитель международного исследовательского проекта «Связанные потребности».

**Alexander Marianovich Abramovich –** M.Sc. in mathematics, visitor researcher (State Key Laboratory of Software Engineering, University of Wuhan, China). Author of more than 50 scientific works. Leader of International research project «Linked Needs».

38/13 Hazionut st., 35312, Haifa, Israel
Тел.: 972-502-1944-36, e-mail: webdao@gmail.com