Development of advisory system based on semantic technologies

Julia Rogushina
Institute of Software Systems
National Academy of Sciences of Ukraine
Kyiv, Ukraine
ladamandraka2010@gmail.com

Anatoly Gladun
International Research and Training Center of ITS of NASU and MESU
Kyiv, Ukraine
glanat@yahoo.com

Serhii Pryima, Oksana Strokan
Dmytro Motorny Tavriya State Agrotechnological University
Melitopol, Ukraine
pryima.serhii@gmail.com,
oksana.strokan@tsatu.edu.ua

Abstract—We analyze main characteristics of modern semantic technologies and consider specifics of their components used for development of the Web-oriented advisory systems. Ontological analysis is used for representation of distributed background knowledge about users, their professions, competencies, lifelong learning outcomes, etc., and standards of the Semantic Web project provide the technological foundation for creation of intelligent advisory applications. Validation of learning outcomes of informal and non-formal learning needs in semantic matching of user profiles with classifications of professions and qualifications.

On base of this research we propose models and means for representation, acquisition and processing of knowledge that make such systems more efficient and dynamic. Methods for comparison of different advisory information objects are based on atomic competences processing are developed. Elements of artificial intelligence and machine learning are used for construction of information objects that are processed by advisory systems.

Advantages of proposed approach are demonstrated on example of applied information system AdvisOnt developed to combine the market of educational services with the labor market that can use validation of the informal and non-formal learning outcomes.

Keywords—ontology, semantic technologies, advisory system

I. INTRODUCTION

The modern world exists in the information space, where a huge number of intelligent information systems (ISs) interact for solving a variety of problems. Over the last years the Semantic Web and associated semantic technologies have moved from the domains of research and standards committees to the mainstream industry. Development of the Web-oriented ISs shows a general trend caused by growth of data volumes: transition from the processing and storage of large amounts of data to the processing and storage of more compact knowledge with a much more complex structure based on various methods, paradigms and technological solutions (such research directions as Big Data, Internet of Things, Web Mining). All of them apply semantization for different steps of information processing and try to take into account meaning of data and relations with domain knowledge.

Ontology-based semantic technologies allow to build powerful applied intelligent applications aimed on analysis and modeling of complex objects and processes of different nature. Development of such intelligent ISs (IISs) is based on the results of knowledge structuring in order to construct the schemas of knowledge bases and to define main subjects and objects of IIS functional support. The intelligence of such ISs is related to automation level of problem solving by use of general and specialized knowledge about user, user problem and adapting the solutions to the current state of the information environment.

One of the most well-known projects of semantic technologies is the Semantic Web [1] aimed at is to transform the Web into global knowledge base. It provides a large number of standards and tools for representation and processing information at the knowledge level. Semantic Web is a powerful instrument for improving the efficiency of distributed and shared access to information and its use by intelligent applications [2]. The main components of the Semantic Web are ontologies [3] for representation of knowledge, Web services [4] for knowledge processing and software agents to represent individual needs of users [5]. This semantic technology is supported by open standards that allow to formalize the semantics of information resources (IRs) and software tools for their search and processing: metadata description language RDF [6]; ontology representation language OWL [7]; query language SPARQL [8] for RDF and OWL.

Another interesting example of semantic technology is OSTIS (Open Semantic Technology for Intelligent Systems) oriented on design of knowledge-driven IISs [9]. The basis of OSTIS technology is the semantic computer code — the standard of semantic representation of information in the IS memory. OSTIS provides integration of heterogeneous applied artificial intelligence systems based on open semantics and unifies knowledge representation, machine learning, creation of algorithms and methods for data processing. It aims formation of intel-
ligent environments for human life (Internet of Things, smart home, smart highways, Smart City, etc.). This semantic technology can be used in design of various ISs: intelligent control systems, reference (consulting, advisory) systems, interactive textbooks and training systems, processing of the Web of Things, etc.

Use of knowledge representation and processing methods depends on specifics of solved tasks. In addition, domain of task defines sources of domain knowledge and characteristics of these sources. Semantic applications [10] are a subset of intelligent ISs. In these work we analyze the specifics of semantic ISs, their components and features and demonstrate advantages of semantic technologies on example of he Web-oriented advisory system AdvisOnt.

For this purpose we consider tasks dealt with combining of the market of educational services with the labor market that can solve problems caused by informal and non-formal learning. Therefore we take into account specifics of educational domain, its subjects and objects and their relations into the labor market. We consider open sources of information about them and methods of their acquisitions and matching.

II. PROBLEM DEFINITION

We analyse main elements of semantic technologies aimed on more efficient development of the Web-oriented intelligent applications. Ontology-based approach provides a lot of advantages in development of applied software but it makes calculations much more complex and cumbersome. Therefore every practical task needs in analysis of expedience of all components of semantic technologies.

This problem we consider on example of advisory system AdvisOnt that provides automated semantic methods for matching of qualifications and competences of various information objects (IOs) – humans, organizations, learning courses, requirements of employer etc. Ontology of competencies and methods for processing of atomic competences become the ground for integration of such objects described by different terms from various qualification systems. Therefore this system needs in external knowledge sources and in non-trivial methods of their processing, and use of the Semantic Web standards supports these possibilities that can’t be achieved in other ways.

III. LIFELONG LEARNING AND VALIDATION OF ITS RESULTS

Lifelong learning is a key factor in personal and professional development of human [11]. Validation of the results achieved in the process of non-formal and informal learning (knowledge, skills, competencies, etc.) with use of open educational resources is necessary for access to the labor market and lifelong learning [12]. Employer organizations, trade unions, chambers of commerce and industry, national bodies involved in the recognition of professional qualifications, employment services, youth organizations, education providers, and civil society organizations are the most interested parties in providing opportunities for the recognition of non-formal and informal learning.

Official recognition of non-formal and informal learning outcomes obtained outside formal learning systems of partial qualifications has to take into account all outcomes obtained by persons in process of lifelong learning. Information about these outcomes can be proposed by person or be acquired from various external sources. The procedure of recognition consists of the following mandatory steps:

1) submission of documents that directly (such as educational declaration) or indirectly (such as reports, articles, descriptions of projects etc.) certify learning outcomes of person;
2) creation of group that can determine the possibility of forms and terms of attestation for recognition of learning outcomes acquired in non-formal and informal education;
3) attestation of these learning outcomes recognition.

After validation and formalization oriented on recognition of the results of previous formal education such outcomes are integrated with results of formal education to transform all competencies of person into some interoperable representation. All stages of this process need in external knowledge about learning domain that helps to define meaningful relations between elements of learning outcomes.

Learning outcomes of some person consist of his/her knowledge, skills, abilities and competencies. Recognition of such outcomes that is achieved through non-formal and informal (spontaneous) learning with the help of various open educational resources is necessary for semantic matching of resumes with vacancies of labor market and propositions of learning organizations.

Rapid expansion of the information technologies, data storage and software for the analysis of Big Data and the Web resources sets fundamentally change the ways of information interchange on the labor market. Subjects of the labor market have possibilities to describe their proposals and requirements through various informal characteristics which are often intangible (e.g. team spirit, social skills, leadership skills). Use of different terms to describe such characteristics actualizes the problem of comparing the semantics of such descriptions.

Validation of learning outcomes is aimed on recognition of learning outcomes that involves confirmation by the competent authority of the fact that such outcomes obtained by a person are evaluated according to certain criteria and meet the requirements of the standard. Validation allows the recognition of learning outcomes...
obtained outside the institutions of formal learning (in non-formal and informal education) and is necessary for access to the labor market and lifelong learning [13].

The tools used in this process have to take into account changes in the open world and be dynamic, and they need in semantic retrieval components based on ontological models of user and domain [14].

IV. Semantic Technologies

Semantic technologies that are aimed at knowledge-level processing of information can solve this problem because they are able to formalize, analyze and process the content of information resources (IRs). Now these technologies based on the analysis of domain knowledge and personal information about users are widely used in development of various distributed applications.

Semantics allows to define explicitly meanings and relations between domain concepts represented by data (words, phrases, symbols, etc.) that depends on context. For the same piece of information, semantics can be defined differently depending on ontology used to formalize the user’s view of the world. Ontologies can be used as a formal, explicit specification of conceptualization of terms at a certain level of details.

Semantic processing of information includes: methods and means of integration and unified representation of heterogeneous distributed knowledge and data; retrieval and processing the Web resources as an universal knowledge source of about meaning of information objects.

What technologies and IRs can be named “semantic” at present is an open question that has not common unified definition. Such situation is caused by specifics of domains where various knowledge-based models and methods are applied. Some researchers establish linkage between semantic technologies and processing of natural language, and others – with ontological analysis. In this work we propose to use term “semantic technologies” for such technologies of information processing that distinguish external IRs as sources of knowledge and methods used for automated analysis of this information.

One of the results of such processing is an achievement of semantic compatibility of open educational IRs that allows to use and integrate information about results of non-formal and informal learning from different sources and databases in various ISs.

Semantic technologies in general can be described through a combination of three main components [15]: ontologies; semantic resources; models of semantics of NL-entities.

But this approach is concentrated on processing of natural language (NL) IRs at the semantic level. In our study we take into account more wide classes of IOs with various structure defined by appropriate ontologies. For example, advisory systems analyze people, organizations, vacancies, learning courses, etc. that can contain NL definitions, multimedia elements and structured data.

Therefore from the point of view of creating semantic applications, these three basic components form a hierarchy where:

- ontologies and other knowledge bases (KBs) are the upper level of abstraction of the knowledge structure;
- IO models represent the intermediary level that allows to distinguish typical IOs and their properties and characteristics;
- semantic IRs are the lower level that provides information about individuals of classes.

Semantic IRs can have links between content elements (IOs of various types and structure) and with elements of IO models (for example, links with other Wiki pages or with data). Meaning of links is provided by means specific for IR representation and markup. For example, Semantic MediaWiki uses semantic properties.

Structure of IO models can contain relations with other IOs (for example, some IO of category “Person” has semantic link with IO of category “Organizations” by relation “Place of work”, and such link can be used into page content only if is present into the IO model). IO models can be formalized by various representations such as templates and forms.

Domain ontologies contain classes and individuals of concepts and formalize their properties and characteristics. Hierarchy of components used by semantic applications is demonstrated by Fig. 1.

![Figure 1. Hierarchy of components of semantic applications.](image-url)
learning, the European Commission has developed outcomes in both formal and non-formal and informal ([18], [19]) are analyzed by many researchers. Advantages of the use of ontologies for semantic markup of various documents (NL texts, multimedia context etc.) in Wiki resources can be semantizied in this way. Advantages of the use of ontologies for semantic structuring as categories and semantic properties. Categories help to link Wiki pages with more general terms and group them, and semantic properties allow to define various semantic features and their values of concept linked with some page. Categories and semantic properties of the Wiki pages can be used as classes and object properties of domain ontology, and names of Wiki-pages – as individuals of ontology. Such domain ontology can be built automatically by special functions of Semantic Media Wiki or by special algorithms according to personal needs of users. Unfortunately, there is no logical or semantic restrictions on ontology building in Semantic Media Wiki. Therefore ontologies provided such possibilities remain the important content of semantic technologies as a source of domain knowledge.

V. ONTOLOGIES IN SEMANTIC TECHNOLOGIES

Ontological analysis is widely used now for formal modeling of various domains [16]. Ontology provides a formal explicit description of domain concepts (classes and individuals), their properties, attributes and relations. Moreover, ontology can contain some domain-specific restrictions on use of all these elements and their combinations. Modern intelligent applications use ontologies as interoperable knowledge bases [17], where information acquired by one ISS can be reused by other one.

The most high-usage languages of ontology representation are dialects of OWL (Ontology Web Language) that differ by their complexity and expressiveness: from OWL Lite to OWL Full. The most popular of them is OWL-DL based on descriptive logics Ontology built on OWL is a sequence of axioms and facts with the addition of references to other ontologies associated with it.

A lot of oriented on the Semantic Web software systems use ontologies as a base of domain knowledge for semantic markup of various documents (NL texts, Wiki resources, other semi-structures and structured texts, multimedia context etc.). Popular representation of information in Wiki resources can be semantizied in this way. Advantages of the use of ontologies for semantic representation for learning domain and competencies ([18], [19]) are analyzed by many researchers.

In order to support the process of validation of learning outcomes in both formal and non-formal and informal learning, the European Commission has developed a free internet portal for multilingual classifier ESCO (European Skills, Competencies, Qualifications and Occupations) [20] that joins the labor markets of the EU member states and allows jobseekers and employers to communicate more effectively with definitions of skills, training and work in all European languages. The main elements of ESCO are professions, skills and qualifications related to the labor, education and training market in the EU. ESCO allows users to determine what knowledge and skills are usually required to work in a particular profession. Each ESCO concept is associated with at least one term in all ESCO languages. Thus, ESCO is a source of information on competencies relevant to the labor market in the international dimension, both for the development of higher education standards and for the review of educational programs in higher education, given that professional standards are currently lacking in many professions. ESCO is published as Linked Open Data, and developers can use RDF format. In this work we consider ESCO as source ontology for semantic application that needs in information about structure of skills and competencies.

VI. IoT AND BIG DATA SOURCES IN SEMANTIC TECHNOLOGIES

Now devices of Internet of Things (IoT) generate a significant amount of structured and unstructured data (Big Data), that can be used in various semantic applications for building of user profiles with the help of based on application of the Semantic Web methodologies and best practices to the IoT data [21].

IoT software implementation for e-learning can be based on wireless (sensor) networks and mobile (portable, portable) computing platforms (smartphones, tablet PCs, embedded computing modules). For example, IoT application for the use of smart cards based on NXP Semiconductors’ MAD Standard - MIFARE Application Directory can be used in e-learning systems, Big Data storage, ensuring the security of this data, personal or group user authentication [22].

The use of smart cards has increased the level of validation of non-formal and informal training for home office employees to improve their competencies during the quarantine period.

VII. ADVISONT SYSTEM

Agricultural advisory systems are widely used now for fast dissemination of agricultural knowledge and information, introduction of modern scientific research and technologies in production, mobility and constant advanced training of agricultural specialists. Their implementation becomes an important factor in competitiveness of rural economy.

Development of the agricultural sector causes the dissemination of modern knowledge among agricultural
manufacturers, relevant and efficient training and information support of their employees. AdvisOnt is an agro-advisory system that ensures consulting services for the agricultural sector of economy. It implements an ontological representation of advisory knowledge. AdvisOnt provides formalization and harmonization of semantic models of advisory objects with use of semantic identification and documentation of non-formal and informal learning outcomes and competence-based representation of advisory IOs [23].

We consider this system because in analysis of semantic application we need in information about realization of knowledge base and methods of its processing. Therefore we can propose objective information about compliance to the Semantic Web requirements only for those software where we participate in its design and development.

A. General architecture of AdvisOnt

The general architecture of AdvisOnt defines relations between main subjects of advisory activities (Fig. 2):

- applicant — person needed in some work in agricultural domain and has a set of relevant competencies and skills;
- employer — person or organization needed in employees for execution of some task or work on some position;
- providers of learning services — organizations that propose various (formal, non-formal and informal) learning means for expansion of personal competencies;
- advisors — experts specialized in agricultural domain of fixer region that can use domain knowledge for refinement of mutual interests of employers and applicants and provides advising services if applicant qualification needs in additional learning according to employer demands.

AdvisOnt helps in interaction between expert-advisor and other subjects by e-Extension interface and uses external semantic IRs and knowledge bases: ESCO as a source of structured representation of domains competencies and qualifications; user profile ontology to determine the structure of the applicant’s model; domain ontologies containing facts and rules of specific agricultural tasks; expert knowledge and soft skills used for semantic formalization and matching of vacancies and resumes; ontology of open online learning services (such as Massive Open Online Courses (MOOCs)) [24].

All classes of ESCO ontology used by AdvisOnt are stored into Turtle file. SPARQL queries and connectors are used for selection of skills and occupations from this RDF repository. The answers of SPARQL queries can be represented as result sets or RDF graphs. In the same way, the results of requests are returned to the RDF repository. Analysis of this ontology is used to define semantic similarity estimates for competence concepts [25].

Domain ontologies are integrated into the RDF repository with use of database of semantic graphs GraphDB. This database complies with W3C standards and links data from various sources, indexes them for semantic search and uses elements of NL analysis. GraphDB connectors provide fast search for keywords and aggregations usually realized by external services with use of synchronization on level of entities defined by URI, properties and property values.

AdvisOnt can use IoT information about user informal competencies and skills acquired from his/her mobile devices to estimate time and quality of learning. Processing of Big Data is not realized now but in future AdvisOnt can import domain rules resulted from Big Data and their metadata by other semantic applications and represented on base of ontologies.

We can consider AdvisOnt as a semantic application because it provides [26]:

- personified interaction for potential employers and job seekers based on use of personal intelligent agents;
- registration of vacancies and resumes with their semantic analysis for formalized representation of used terms (on base of NL texts markup by ESCO ontology concepts);
- comparison of resumes and vacancies at the semantic level with use of semantic similarity of domain terms semantic relations between professions, knowledge, skills, competencies and qualifications defined by ESCO ontology;
- personified search of educational services and training courses based on validation of learning outcomes of formal, informal and non-formal level;
- comparison of training courses and programs with professions on base of their atomic competencies [22].

B. Semantic components of AdvisOnt

From the point of view of relations between components of semantic technology we can AdvisOnt contains:

- external KBs represented by ESCO ontology, MOOC ontology for learning courses, various agriculture domain ontologies and user profile ontologies from other IISs and internal ontology of competencies and qualifications;
- IO models that formalize structure and features of typical advisory IOs: competencies, skills and professions, applicants, vacancies and resumes, etc.;
- IRs that contain semantic markup based on structure of typical IOs and provide additional information about individuals of classes: semantic Wiki resources that can contain NL text and multimedia (Fig. 3).
C. AdvisOnt as a Semantic Web application

We can rate AdvisOnt as semantic application of the because this IS conforms to the requirements of the Semantic Web Challenge: Minimum requirements of the Semantic Web application for AdvisOnt can be interpreted like this:

1) data meaning plays a key role in its functioning: AdvisOnt process meaning of vacancies and resumes with use of ontologies to link various terms with concepts and realize original non-trivial approach based on atomic competencies for matching of IOs that cannot be obtained without analysis of their semantics.

2) AdvisOnt uses ontologies and IRs from different owners (ESCO, MOOCs, etc.) that can be hanged by other ones (for example, by ontology of national qualification system or other e-learning platform) without changes of software, these sources are heterogeneous syntactically (ontologies, Wiki IRs, thesauri, etc.) and semantically (use different NL languages and describe various domains), and contain real-world data used by other commercial applications.

3) Search for information is carried out in the real information space of the Web: results of AdvisOnt depend on user requests and actual information retrieved from the Web about vacancies, resumes and learning courses.

AdvisOnt works into the open information space, i.e.
recommendations are not absolutely optimal but are based on available data and knowledge of system. This IS is based on processing of IRs that are represented on languages developed by the Semantic Web — RDF and OWL.

VIII. CONCLUSION AND FUTURE WORK

We consider semantic approach to development of advisory systems that provides methods and means of integration and unified representation of heterogeneous distributed knowledge and data. Analysis of the Web semantic resources as an universal knowledge source of about meaning of information objects helps in selection of external information sources of advisory system.

Analysis of the development of modern ITs based on knowledge shows the feasibility of using the Semantic Web initiative in the development of applied IS for validation of the on formal and non-formal learning outcomes that can to combine the market of educational services with the labor market. Conformity of developed IS with requirements to the Semantic Web application provides more flexible, intelligent and personified IS that is oriented on processing into the open information space and use of interoperable ontological knowledge.

In future we plan to develop more detailed models of advisory IOs and use II methods for acquisition of knowledge about these IOs from various Web IRs (such as Big Data and IoT).

REFERENCES

Разработка консультирующих систем на основе семантических технологий

Рогушина Ю.В., Гладун А., Прийма С., Строкань О.

Анализируются основные характеристики современных семантических технологий, рассматриваются их компоненты, которые используются для создания веб-ориентированных консультирующих систем. Онтологический анализ используется как основа для формирования фоновых знаний о пользователях, их профессиях и компетенциях.

Стандарты Semantic Web обеспечивают технологический базис создания интеллектуальных приложений, поддерживающих процесс консультирования. Валидация результатов неформального обучения требует сопоставления профилей пользователей с классификаторами профессий и компетенций. На основе проведенного исследования предлагается модели и методы представления, извлечения и обработки знаний, которые обеспечивают эффективность и динамичность разрабатываемой системы. Методы сопоставления информационных объектов консультирования базируются на сравнении наборов атомарных компетенций.

Received 24.05.2021