ANNEX 3. Abstracts of Publications, submitted by Assoc. Prof. Gennady Agre to the competition for the academic position "professor" in professional field 4.6 "Information and computer sciences" announced in ДВ №. 41/21.05.2019

Semantic Web Services

1. **Agre, G. INFRAWEBS designer - A graphical tool for designing semantic web services. Lecture Notes in Computer Science 4183 LNCS, 2006, pp. 275-289 (SJR = 0.317)**

In order to make accessible new Semantic Web Services technologies to the end users, the level of tools supporting these technologies should be significantly raised. The paper presents the architecture of such a tool – an INFRAWBS Designer – a graphical ontology-driven tool for creating a semantic Web service description according to WSMO Framework. The tool is oriented to the end users – providers of Web services, who would like to convert their services into WSMO based semantic Web services. The most character features of the tool – intensive use of ontologies, automatic generation of logical description of a semantic service from graphical models and the use of similarity-based reasoning for finding similar service descriptions to be reused as initial templates for designing new services are discussed. By using the INFRAWEB Designer, the user can create a correct logical description of a WSMO based semantic Web service without knowing WSML - the formal language for describing such services.

2. **Agre, G., Marinova, Z., Pariente, T., Micsik, A. Towards semantic Web service engineering. CEUR Workshop Proceedings 243, 2007, pp. 91-105 (SJR = 0.157).**

The paper presents the main results of the IST FP6 INFRAWEBS project. The project has developed an easy and effective way of constructing and using semantic descriptions for existing and new Web services. INFRAWEBS has adopted the WSMO (Web Service Modeling Ontology) and WSML (Web Service Modeling Language) specifications and has imposed no additional requirements to them. Therefore, the advanced software components developed during the project are of interest to the whole WSMO community. INFRAWEBS offers a SOA framework – INFRAWEBS Integrated Framework (IIF), based on ESB integration paradigm, which can be easily used by different groups of users (application providers, designers of semantic Web services, etc.). The IIF enables integration of components implemented by means of different program technologies. Furthermore, it can be considered as one of the first frameworks for semantic service engineering that covers the whole SWS life-cycle and allows for creation of complex, semantically-enabled applications.

3. **Agre, G., Dilov, I. How to create a WSMO-based semantic service without knowing WSML. Lecture Notes in Computer Science 4832 LNCS, 2007, pp. 217-235 (SJR = 0.293)**

INFRAWEBS Designer is a graphical, ontology-based, integrated development environment for designing WSMO-based semantic Web services and goals. The INFRAWEBS Designer is oriented to Web service providers and Web service application providers, and does not require any preliminary knowledge of WSML – the logic language for creating semantic service description. The Designer allows a user to graphically compose a WSMO-based semantic description of a given Web service based on existing WSDL description of this service and a set of WSML ontologies. To facilitate this process the graphical descriptions of

look-alike semantic Web services can be found and re-used. Such descriptions can serve as templates for the WSMO object under construction. The main objective of this paper is to show how this tool can be used by service providers to create a WSMO-based semantic Web services and to present some conclusions on the tool applicability based on our experience in creating two pilot INFRAWEBS applications - the first uses a travel agency scenario and the second is based on an eGovernment scenario.

4. **Agre, G; Lobo, TP; Marinova, Z; Nern, J., Micsik, A., Boyanov, A., Atanasova, T., Scicluna, J., Lopez-Cobo, J.M, Tzafestas, E. Chapter 12: INFRAWEBS - A Framework for Semantic Service Engineering. In: DiNitto, E; Sassen, AM; Traverso, P; Zwegers, A (Eds.). At Your Service: Service-Oriented Computing from an EU Perspective, 2009, 299-325, M I T PRESS, USA (WoS)**

The chapter presents main results of the IST research project INFRAWEBS, which focused on developing a semantic service engineering framework enabling creation, maintenance, and execution of WSMO-based Semantic Web Services, and supporting semantic Web Service applications within their life cycle. Being strongly conformant to the current specification of various elements of WSMO, the INFRAWEBS framework manages the complex process of creation of semantic descriptions by identifying different types of actors (users) of Semantic Web Service technologies; clarifying different phases of the semantic service engineering process; and providing a specialized software tool set oriented to the identified user types and intended for usage in all phases of the Semantic Web Service engineering process. The viability of the INFRAWEBS approach has been tested by two pilot applications with a high level of complexity. The first application is the STREAM Flows! System, in which customers can create and reuse travel packages. The second application is based on an e-government scenario. It illustrates interactions carried out by Semantic Web Services in the scope of public administration and interactions between administrations, citizens, and enterprises, with emphasis on the enterprise-to-administration integration. The set of tools developed during the project makes the otherwise arduous process of constructing a WSMO-based Semantic Web Service application, requiring expert knowledge of the WSMO model, and the Web Service Modeling Language (WSML), a task achievable by ordinary Web Service developers. Facilitating the task of WSMO object construction is an achievement of the project that has a potential impact on the adoption of Semantic Web Services on a larger scale. The set of software modules and tools developed during the INFRAWEBS project provides some innovative solutions for semantic service description, publishing, discovery, and dynamic composition, with the main focus on solving problems occurring in the process of creating real service-based applications. The INFRAWEBS framework can be considered as the first framework for semantic service engineering that covers the whole Semantic Web Service life cycle and allows creation of complex semantically enabled applications.

5. **G. Agre. Chapter 5: Engineering Semantic Web Services. In: Gabriel Fung (Ed .). Engineering Semantic Web Services. 2010, iConcept Press, USA, pp. 75-98, ISBN: 978-0-9807330-1-3 (Hard Cover) / 978-1-4536364-0-4 (Paperback)**

Semantic Web services (SWS) research is related to automating the development of Web service based applications through semantic Web technology. By providing formal descriptions with well defined semantics, SWS are another step in the direction of solving service engineering problems such as service inter-operation, discovery, choreography and orchestration. The IST FP6 research project INFRAWEBS successfully completed in the beginning of 2007, proposed a technology-oriented step for overcoming some of the above-

mentioned problems. It focused on developing a Semantic Service Engineering Framework enabling creation, maintenance and execution of WSMO-based SWS, and on supporting SWS applications within their life cycle. The present chapter focuses mainly on two hot topics of SWS Engineering – creating SWS description and dynamic composition of services and describes in details original solutions for these problems developed by INFRAWEBS project. Description of SWS is done by means of the IFRAWEBS Designer - a graphical, ontology-based, integrated development environment for designing WSMO-based SWS and goals. The INFRAWEBS Designer is oriented to Web service providers and Web service application providers, and does not require any preliminary knowledge of WSML. The dynamic composition of semantic Web services is based on an original data-driven approach, in which the process of finding an appropriate service composition is guided by the algorithm for run-time decomposition of the user goal into sub-goals and discovering the existing services able to satisfy them. Compatibility of services participating in the composition is achieved by using the consistent description of the composite goal template, prepared by the service application provider in design-time. Using only implicitly provided information about the desired order of execution of services in the composition, the proposed algorithm is able to find a proper orchestration of services in the composition as well as to discover the appropriate service substitutions when some of the services in the composition can not be executed due to some physical reasons.

Technology Enhanced Learning and Semantic Technologies

6. **Dochev, D., Agre, G. Towards semantic web enhanced learning. KMIS 2009 - 1st International Conference on Knowledge Management and Information Sharing, Proceedings, 2009, pp. 212-217 (Scopus)**

Learning content has always been regarded as a keystone for all learning situations in classical education as well as in e-learning activities. That's why, the authoring of learning materials is one of the most important and labour-intensive activities in the modern Technology Enhanced Learning (TEL) practice. In this paper we have proposed an approach for developing a new Semantic SOA-based framework oriented to TEL applications facilitating reusability and repurposing of learning objects. The approach is based on analyzing and exploiting the advantages of SWS technology in the automation of learning object discovery, selection and composition within a distributed service architecture seamlessly integrated through ontologies. The approach reflects current work in the frame of an ongoing Bulgarian research project SINUS "Semantic Technologies for Web Services and Technology Enhanced Learning".

7. **Dochev, D., Agre, G., Pavlov, R. An approach to learning-by doing through user creation of learning content. EUROMEDIA 2011 - 16th Annual Scientific Conference on Web Technology, New Media Communications and Telematics Theory Methods, Tools and Applications, 2011, pp. 9-12 (Scopus)**

The paper investigates the organization of learning-by-doing activities through learner's authoring of analytical materials, considering some specifics of education in humanities. Semantic technologies are applied to guide and help the learner's actions in developing limited-sized dedicated collections of multimedia objects, adequate to pre-assigned learning tasks and then in comparing the selected objects for the needs of performed analysis. The discussed framework for a TEL environment targets a concrete humanitarian domain - Bulgarian Iconography with educational uses in a set of disciplines like iconography, arts,

history, culture studies, theology, etc. The paper contains an example with structured formulation of a given learning task, and its formalization in form of queries to the environment to help in the collection development steps and in the evaluation of the adequacy of the selected by the learner representative subset of objects.

8. **Dochev, D., Agre, G., Pavlov, R. User authoring in learning-by-doing situations. ACM International Conference Proceeding Series 578, 2011, pp. 577-582 (SJR = 0.184)**

The paper deals with learning-by-doing activities through learner's authoring of analytical materials in well-defined learning situations. These activities are facilitated by Semantic Web techniques to support the learners in the access and filtration of necessary information objects to be analyzed during the authoring process as well as (to some extent) in the evaluation of the created materials. The paper discusses an experimental application of a Technology Enhanced Learning environment that under development in the frame of the current Bulgarian NSF project SINUS "Semantic Technologies for Web Services and Technology Enhanced Learning" to the task for creating a collection of Bulgarian icons satisfying some preliminary specified requirements.

9. **Hristov, I., Agre, G., Dochev, D. Measuring the learning progress in a "learning by authoring" semantic web services based ecosystem. CSEDU 2011 - Proceedings of the 3rd International Conference on Computer Supported Education 2. 2011. pp. 413-418 (Scopus)**

The paper describes an approach for helping the evaluation of specific learning-by-authoring activities, producing analytical essays/projects on limited-sized dedicated collections of multimedia objects, created by the learners according to pre-defined learning tasks. The three types of learner activities are continuously evaluated for giving the support to the learner. The activities related to collection creation are evaluated on the bases on the comparison of the results of learner's queries and formalized learning task descriptions prepared by the instructor. The evaluation of quality of the analysis is made by the evaluator. The evaluator is supported by the system with measurable counters of concepts used or missed by the learners and the percentage of conjunction in concepts. The visual appearance, which, in our case, is not subject of learning, is rated by the other participations in the project.

10. **Dochev, D., Agre, G. Supporting learning-by-doing situations by semantic technologies. EUROMEDIA 2012 - 17th Annual Scientific Conference on Web Technology, New Media Communications and Telematics Theory Methods, Tools and Applications, 2012, pp. 49-53 (Scopus)**

The paper discusses an approach to learning-by-doing activities through learners' authoring of analytical materials in specific learning settings and presents an experimental learning environment that applies this approach by using semantic information technologies. It is briefly presented how the environment may guide and consult the learners, considering the normal shortage, inaccuracy and even incorrectness of the initial learners' knowledge for the domain and also for the accessible materials and available information support. The discussed environment is experimented in a concrete domain - Bulgarian Iconography with educational uses in a set of humanitarian disciplines. An example of how the environment may assit the learners in execution of some concrete learning tasks - the development of a dedicated multimedia collection and its analysis and evaluation are presented.

11. Staykova, K., Agre, G. Use of Ontology-to-Text Relation for creating semantic annotation. ACM International Conference Proceeding Series, 2012, pp. 64-71 (SJR = 0.181)

The paper is focused on the problem for creating new semantic annotations of multimedia objects stored in a digital library. In most cases creating such annotations is a time-consuming and expensive process, when it is carried out manually by domain experts. This cumbersome process could be facilitated if a part of the information used for annotating is available in form of natural language texts. In such cases it is possible to apply some Natural Language Processing techniques for extracting needed information and then convert it into desired semantic annotations. The paper describes an approach for solving the task of ontological terms recognition within Bulgarian texts by application of a Natural Language Processing technique called "Ontology-to-Text Relation". All preliminary steps to prepare needed resources for implementation of this technique are discussed.

12. Agre, G., Dochev, D., Slavkova, L. Technology Enhanced Learning for humanities by active learning - The SINUS project approach. Cybernetics and Information Technologies. 2012, 12(4), pp. 25-42 (SJR = 0.111)

The present paper deals with technology Enhanced Learning (TEL) support for specific active learning activities − learner's authoring of analytical materials by intensive use of digital content from multimedia Digital Libraries (DL). In order to support the learners, the necessary types and content of explicit built-in domain knowledge were identified and formalized. The proposed approach to semantic modelling of the domain knowledge consists in combining a basic ontology, describing only features, explicitly or implicitly built in the structure of the used DL, and additional specialized ontologies. This approach allows keeping intact the content, annotation and access method of the DL used and at the same time it enables the enrichment of the semantic access to DL information objects by additional descriptive (ontological) features through attachment of specialized ontologies. The focus of the analysis was put on the learner's actions during the first phase of the defined learning task – the development of a dedicated collection of DL multimedia objects. Though seeming simpler and easier to be formalized, this phase has a potential for different levels of non-prescriptive active learner's behaviour in the three possible modes of operation with the designed TEL environment. The proposed models of correct solutions and of possible learner's errors shape the developed error detecting algorithm, checking the correctness and completeness of the learner's solution irrespective of the way he/she has obtained it. These design decisions shaped the functionality of the experimental SINUS TEL environment, implemented using up-to-date semantic technologies.

13. Agre, G. SINUS - A Semantic Technology Enhanced Environment for Learning in Humanities. Cybernetics and Information Technologies. 2012, 12(4), pp. 5-24 (SJR = 0.111)

The paper describes a SINUS environment − a semantic technology based environment intended for developing technology enhanced learning applications in humanitarian problem domains. The environment consists of three layers: the storage layer contains heterogeneous repositories storing domain and pedagogical knowledge; the tool level contains a set of tools for processing different types of knowledge, and the middleware layer that is implemented as an extended search engine carrying out all necessary communications between the tools and

the repositories. The environment has a high degree of heterogeneousness: at the moment it comprises three types of repositories – a relation database used for storing learning task models, a WSDL Web service implementing all functions of an SQL-based basic digital library storing multimedia objects and their annotations, and a semantic RDF-based repository storing OWL ontologies and semantic annotations of objects described in the basic digital library. The environment components are written in different programming languages such as C# (Extended Search Engine) and Java (Semantic Annotation Editor and Learning Task Editor) that are implemented as WSDL or RESTful Web services. Some implementation issues of the successful use of such heterogeneous tools for creating the exemplary TEL applications (the extended digital library with the service oriented architecture and the learning system in the domain of the East Christian iconography) are discussed. A preliminary evaluation of the environment that has proved the correctness of the chosen approach for developing it, is also presented.

14. Dichev, C., Dicheva, D., Agre, G., Angelova, G. Current practices, trends and challenges in k-12 online learning. Cybernetics and Information Technologies. 2013, 13(3), pp. 91-110 (SJR = 0.173)

Online learning, a general systemic approach for education that uses a new delivery medium − Internet, is one of the fastest growing trends in Technology Enhanced Learning. Technology in combination with an instruction that addresses the cognitive and social processes of knowledge construction could offer more diverse and effective online learning opportunities than their face-to-face counterparts. In this review we attempt to summarize different forms and practices in primary and secondary education (K-12) online and blended learning as they appear in various regional, national and cultural contexts. The article starts with introducing some basic concepts and terminology, sums up the state of K-12 online learning around the world and ends with summarizing some trends and challenges observed in current K-12 online learning practices.

15. Dichev, C., Dicheva, D., Angelova, G., Agre, G. From gamification to gameful design and gameful experience in learning. Cybernetics and Information Technologies. 2014, 14(4), pp. 80-100 (SJR = 0.138)

Learning is a goal driven social activity determined by motivational factors. To be able to efficiently gamify learning for improved student motivation and engagement, the educators have to understand the related aspects studied in games, motivational psychology and pedagogy. This will help them to identify the factors that drive and explain desired learning behaviors. This paper presents a survey of the main approaches employed in gamification and the emerging new directions in the context of the relevant motivational psychology and pedagogy. The focus is on the motivational factors that impact learning and understanding of behavior change. The purpose of the paper is two-fold: on one side, to provide analysis and guide to relevant works related to gamification along with outlining the emerging trends, and on the other, to provide foundation for evaluation and identification of the areas of possible improvements.

16. Dicheva, D., Dichev, C., Agre, G., Angelova, G. Gamification in education: A systematic mapping study. Educational Technology and Society, 2015, 18(3), pp. 75-88 (IF – Q2)

While gamification is gaining ground in business, marketing, corporate management, and wellness initiatives, its application in education is still an emerging trend. This article presents a study of the published empirical research on the application of gamification to education. The study is limited to papers that discuss explicitly the effects of using game elements in specific educational contexts. It employs a systematic mapping design. Accordingly, a categorical structure for classifying the research results is proposed based on the extracted topics discussed in the reviewed papers. The categories include gamification design principles, game mechanics, context of applying gamification (type of application, educational level, and academic subject), implementation, and evaluation. By mapping the published works to the classification criteria and analyzing them, the study highlights the directions of the currently conducted empirical research on applying gamification to education. It also indicates some major obstacles and needs, such as the need for proper technological support, for controlled studies demonstrating reliable positive or negative results of using specific game elements in particular educational contexts, etc. Although most of the reviewed papers report promising results, more substantial empirical research is needed to determine whether both extrinsic and intrinsic motivation of the learners can be influenced by gamification.

17. **Dichev, C., Dicheva, D., Agre, G., Angelova, G. Trends and opportunities in computer science OER development. Cybernetics and Information Technologies, 2015, 15(3), pp. 114-126 (SJR = 0.158)**

The world is embracing an open education model. The success of this process implies an adequate awareness, an assumption that is inconsistent with recent reports and statistical facts. Despite the major advances in recent years, Open Educational Resources (OER) are still not in the mainstream of Computer Science course development. Motivated by the need to fill this gap, this paper analyzes the evolution of the OER development and the emerging trends relevant to Computer Science education. The aim is to raise the awareness and promote a practical transition process towards an adequate model that aligns the interests of all stakeholders. In order for CS OER to deliver on its potential to improve the CS education, the sustainable business models, increased awareness, relevant policies and accreditation systems are among the key issues to be overcome.

18. **Dobreva, M., Angelova, G., Agre, G. Bridging the gap between digital libraries and e-learning. Cybernetics and Information Technologies, 2015, 15(4), pp. 92-110 (SJR = 0.158)**

Digital Libraries are offering access to a vast amount of digital content, relevant to practically all domains of human knowledge, which makes it suitable to enhance teaching and learning. This paper presented an overview of the current approaches to close the chasm between the potential usefulness of digital libraries in education and the eLearning environments currently in use. In practical terms we believe that the areas of highest importance and urgency are to achieve a clear understanding of needs in content in eLearning, and to release more open educational resources. The better understanding of needs – not as perceived by the digital library community but as experienced in the educational community, will help to bridge the gap between both communities. In addition, offering more open educational resources would be a helpful strategic step for strengthening the engagement of digital libraries with education. We hope that the systematic efforts in these directions will bring a positive change and the wealth of digital libraries' resources and will be of better service in education for the benefit of all teachers and learners.

Machine Learning and Data Mining

19. Hristov, V., Agre, G. A software system for classification of archaeological artefacts represented by 2D plans. Cybernetics and Information Technologies, 2013, 13(2), pp. 82-96 (SJR = 0.172)

An important part of archaeological analysis is the task for classifying ceramics based on the vessels' shape, their decoration, material, etc. The shape is one of the fundamental properties of a vessel used for its identification. However, the traditional shape descriptions rely on intuitive, often vague characterizations, which are hard to quantify, or based on some easily measurable attributes (like vessel's height, diameter of its mouth, etc.), describing the vessel's shape only partially. The present paper describes some initial results related to the ongoing project aiming at developing a system facilitating archaeological analyses of artifacts based on their representation as 2D archaeological drawings. The emphasis is on the task of identification (classification) of ceramic vessels (or shards) which belong to a discrete set of preliminary defined types (or classes). The input information for solving this task is a scanned drawing of pottery (from a book or an archaeological report) and an Internet-based database containing preliminary classified instances of vessels belonging to some vessel types. The method for representing whole and fragmented ceramic vessels as well as process of construction of such a representation is described. Some initial results of experiments, using the proposed representation for solving the task of identification of ancient Greek amphorae found in the closed archaeological complexes on the territory of the Black sea coast are discussed.

20. Strandjev, B., Agre, G. On applicability of Principal Component Analysis to concept learning from images. 2013 IEEE International Symposium on Innovations in Intelligent Systems and Applications, IEEE INISTA 2013, IEEE Xplore, August, 2013, DOI: 10.1109/INISTA.2013.6577623 (Scopus)

The paper presents some experiments investigating the applicability of the Principal Component Analysis method for solving several concept learning tasks defined on images of faces. The results have shown that, in most cases, the applied transformation improves the classification accuracy of concept learning algorithms used. In addition, the experiments have confirmed a possible relation between the quality of the obtained improvements and the complexity of the concepts to be learnt. This relation has the potential to be an objective measure of "concept complexity" notion.

21. Strandjev, B., Agre, G. On impact of PCA for solving classification tasks defined on facial images. International Journal of Reasoning-based Intelligent Systems. 2014, 6(3-4), pp. 85-92 (SJR = 0.104)

In this paper we have discussed the impact of the principle component analysis (PCA) for solving tasks related to classification of different abstract concepts learnt from facial images. The conducted experiments have shown that in most cases when the concept to be learnt is reasonable (i.e., not artificially defined) the application of PCA leads to improvement in classification accuracy of the nearest neighbour classifier used to recognise such a concept. Additionally a possible approach for measuring the complexity of the concept to be learnt has been briefly discussed.

22. Marinchev, I., Agre, G. A customised metric for foods categorization. ACM International Conference Proceeding Series 883, 2014, pp. 234-239 (SJR = 0.25)

The paper presents a heuristic measure specifically tailored toward categorization of food products, which is a part of the ongoing project with the National Innovation Fund for developing expert system for dietary and healthy nutrition. The measure is based on using two functions – the first evaluates the distance between the food product to be categorized and the etalon instances of the food categories represented by a food ontology; and the second measures the level of trust that we can have to the solution implied by the found nearest category instance. All foods are described by a set of attributes that corresponds to their nutrients. Such a representation is the appropriate representation of foods data since the content of the food is the base for developing healthy menus and plays a critical role in the support of the physical fitness and personal appearance. Several experiments have been performed demonstrating that the approach leads to the results, which are good enough when more general categories (categories from the upper levels of the food ontology) are in use or when the task for finding the appropriate substitutes of a given product should be solved.

23. Marinchev, I., Agre, G. On speeding up the implementation of nearest neighbour search and classification. ACM International Conference Proceeding Series 1008, 2015, pp. 207-213 (SJR = 0.253)

The paper presents some practical approaches and techniques to speeding up implementations of the nearest neighbour search/classification algorithm in the case of high dimensional data and/or many training examples. Such settings often appear in the fields of big data and data mining. We apply a fast iterative form of polar ortogonalization and use the computed decomposition matrix to pre-select smaller number of candidate classes for the query element. It has been shown that additional speed up can be achieved in the case of great number of training examples per class by subdividing them into subclasses by fast approximation of some clustering algorithms and by using new classes for building the decomposition matrix. The proposed pre-processing step, which depends linearly or near linearly on the number of examples and dimensions, as well as the pre-selection step depending on the number of classes can be used with every well-known indexing method (such as annulus method, kd-trees, metric trees, rtrees, cover trees, etc.) to limit the number of training examples used in the search/classification process. Finally we propose a so called “cluster index” structure that practically extends the applicability of the indexing structures with higher order complexity to bigger datasets.

24. Marinchev, I., Agre, G. An expert system for healthful and dietary nutrition. ACM International Conference Proceeding Series 1164, 2016, pp. 229-236 (Scopus)

The paper presents an expert system for healthful and dietary nutrition developed as part of a research project financed by Bulgarian SME fund. The main purpose of the system is to facilitate the classification of the unknown food products to the set of predefined classes related to different client groups such as “appropriate or not for children”, “appropriate or not for adults below 60 years with high physical activity” etc. The expert knowledge consists of an ontology of standard food categories describing by nutrients, and by a set of production rules related different food categories with preliminary defined classes. The classification process consists of two steps: the first is the identification of a concrete product by finding a standard product from the ontology that has the biggest similarity with the tested product; and the second step is the application of the appropriate production rules based on the nutrition

content of the product. The paper provides a detailed description of the system architecture, classification process, belief propagation mechanism, different user roles available in the system and as well the system implementation.

25. **Nikolova, I., Dicheva, D., Agre, G., Angelova, G., Dichev, C., Madzharov, D. Emerging applications of educational data mining in Bulgaria: The case of UCHA.SE. Studies in Computational Intelligence 648, 2016, pp. 113-131 (SJR = 0.219)**

This paper presents a study of the educational web portal UCHA.SE aiming to improve the quality of the provided educational services. We employed machine learning algorithms to extract educational data from the system logs that can contribute to the evaluation of the portal and to identifying gaps that can be remediated. By applying subsampling and oversampling methods for manipulating the available training data set with highly imbalanced classes, we constructed a quite accurate and understandable for the end-user rule-based model for predicting whether a user would renew her subscription to the site or not. The model was analyzed in order to find significant factors, which would signal whether a user will continue or discontinue their subscription. The analysis has shown that such important factors are the period of the first subscription, the speed with which the user looses interest to the portal and the completion degree of corresponding exercises after watching videos. The main recommendations to the developer team of the portal are to extend the system with more opportunities for learning about the offered topics, including a variety of exercises and quizzes with different level of difficulty for practicing as well as providing leveled (need driven) help for completing them. This would allow personalization, resource recommendation, and tracking and visualizing learner progress and achievements.

26. **Nikolova, I., Dicheva, D., Agre, G., Angelova, G., Dichev, C., Madzharov, D. Exploring the use of resources in the educational site ucha.SE. Lecture Notes in Computer Science 9883 LNAI, 2016, pp. 347-351 (SJR = 0.339)**

The present paper discusses some results from an on-going pilot project aiming at analyzing and improving the quality of the educational services and respectively the revenue generation for the educational site UCHA.SE. The data used for evaluation was extracted from over 3000000 user accesses to 3797 resources within 65 categories. Each category is described by a set of features constructed on the base of the site‘s system logs and students interactions stored directly in the system database. These features include: (i) total number of video materials in the category; (ii) average number of accesses to the resource category for a 3-months period; (iii) percentage of all students who have accessed the resource category; (iv) number of teachers comments for this resource. Based on these characteristics we perform statistical analysis on the resource availability in the educational site UCHA.SE and the user interest in all 65 resource categories in the portal. The analysis has allowed proposing some possible improvements of the content and helped identifying groups of users, which could be approached in a specific way. The categories which are outliers and need more attention have been pointed out.

27. **Agre, G., Dzhondzhorov, A. A weighted feature selection method for instance-based classification. Lecture Notes in Computer Science 9883 LNAI, 2016, pp. 14-25 (SJR = 0.339)**

The paper presents a new method for selecting features that is suited for the instance-based classification. The selection is based on the ReliefF estimation of the quality of features in the orthogonal feature space obtained after PCA transformation, as well as on the interpretation of these weights as values proportional to the amount of explained concept changes. The user sets a threshold defining what percent of the whole concept variability the selected features should explain and only the first "stronger" features, which combine weights together exceed this threshold, are selected. During the classification phase the selected features are used along with their weights. The experiment results on 12 benchmark databases have shown the advantages of the proposed method in comparison with traditional ReliefF.

28. Agre, G., Petrov, D., Keskinova, S. A new approach to the supervised word sense disambiguation. Lecture Notes in Computer Science 11089 LNAI, 2018, pp. 3-15 (SJR = 0.283)

The paper presents a new supervised approach for solving the all-words sense disambiguation (WSD) task, which allows avoiding the necessity to construct different specialized classifiers for disambiguating different target words. In the core of the approach lies a new interpretation of the notion 'class', which relates each possible meaning of a word to a frequency with which it occurs in some corpora. In such a way all possible senses of different words can be classified in a unified way into a restricted set of classes starting from the most frequent, and ending with the least frequent class. For representing target and context words the approach uses word embeddings and information about their part-of-speech (POS) categories. The experiments have shown that classifiers trained on examples created by means of the approach outperform the standard baselines for measuring the behavior of all-words WSD classifiers.

29. Agre, G., Petrov, D., Keskinova, S. Word Sense Disambiguation Studio: A Flexible System for WSD Feature Extraction. Information, 10(3), 97, Multidisciplinary Digital Publishing Institute, 2019, ISSN:2078-2489Close, DOI:10.3390/info10030097, (SJR = 0.222)

The paper presents a flexible system for extracting features and creating training and test examples for solving the all-words sense disambiguation (WSD) task. The system allows integrating word and sense embeddings as part of an example description. The system possesses two unique features distinguishing it from all similar WSD systems—the ability to construct a special compressed representation for word embeddings and the ability to construct training and test sets of examples with different data granularity. The first feature allows generation of data sets with quite small dimensionality, which can be used for training highly accurate classifiers of different types. The second feature allows generating sets of examples that can be used for training classifiers specialized in disambiguating a concrete word, words belonging to the same part-of-speech (POS) category or all open class words. Intensive experimentation has shown that classifiers trained on examples created by the system outperform the standard baselines for measuring the behaviour of all-words WSD classifiers.