# STUDY MODULE DESCRIPTION FORM

| | | |
|---|---|---|
| Name of the module/subject<br>**Representation of semanties in WEB** | | Code<br>**1010332531010337157** |
| Field of study<br>**Information Engineering** | Profile of study (general academic, practical)<br>**(brak)** | Year /Semester<br>**2 / 3** |
| Elective path/specialty<br>**Information Technologies** | Subject offered in:<br>**Polish** | Course (compulsory, elective)<br>**obligatory** |
| Cycle of study:<br>**Second-cycle studies** | Form of study (full-time,part-time)<br>**full-time** | |
| No. of hours<br>Lecture: **15** Classes: **-** Laboratory: **30** Project/seminars: **-** | | No. of credits<br>**5** |
| Status of the course in the study program (Basic, major, other)<br>**(brak)** | (university-wide, from another field)<br>**(brak)** | |
| Education areas and fields of science and art<br>**technical sciences**<br>**Technical sciences** | | ECTS distribution (number and %)<br>**5 100%**<br>**5 100%** |

## Responsible for subject / lecturer:

dr inż. Andrzej Szwabe
email: Andrzej.Szwabe@put.poznan.pl
tel. 61 665 3958
Faculty of Electrical Engineering
ul. Piotrowo 3A 60-965 Poznań

## Prerequisites in terms of knowledge, skills and social competencies:

| | | |
|---|---|---|
| 1 | **Knowledge** | The student has the knowledge equivalent to first degree studies in the field of Internet technology. |
| 2 | **Skills** | The student has the skills equivalent to first degree studies in the field of Internet technology. |
| 3 | **Social competencies** | The student has the social skills equivalent to first degree studies. |

## Assumptions and objectives of the course:

The main objective of the course is to present means for representing semantics in the Web, with particular emphasis on the Semantic Web and Linked Data technologies that can be used in practice: for automatic or semiautomatic data integration, identification of mappings between semantic data structures / schemes, for identification of functional equivalence of semantically defined web services and for ontology alignment.

The course is designed to prepare the student to design, implement and evaluate a semantic data integration system. Software libraries and software packages are presented that can be used to rapidly create a modern scalable system that is capable of processing and integrating semantic data. In order to enable the student to objectively evaluate a semantic data integration system and conduct its own research in the relevant area, a standard evaluation methodology is presented, along with the key measures of mapping quality and examples of their implementations based on the leading software libraries.

## Study outcomes and reference to the educational results for a field of study

### Knowledge:

1. The student has knowledge of current trends in computer applications and key related problems. - [K_W06]
2. The student has knowledge of the development trends and the most important new developments in information technology. - [K_W14]

### Skills:

1. Student is able - in formulating and solving IT problems - integrate knowledge from different fields and disciplines. - [K_U07]
2. Student is able - by working in a team - build specification fragments of unusual or complex systems. - [K_U08]

### Social competencies:

1. Student is able to think and act in a creative and enterprising way. - [K_K01]

## Assessment methods of study outcomes

Lecture: written exam (evaluation of the theoretical knowledge acquired from the lecture and the ability to design a semantic data integration system for given application scenario).

Laboratory: evaluation of implementations developed during the laboratory exercises (aimed at development of components of a semantic data integration system), the final grade being an average of an average of partial grades (for each class) and a grade of the final implementation of successively prepared components and its presentation in the report.

## Course description

Lecture

The main subject of the course is technology enabling both humans and computer systems to read and write reprezentations of semantic dependencies occurring between objects in the real world. The main purpose of the subject is to present contemporary ways of representing semantics in the Web, with particular emphasis on the Semantic Web and Linked Data technologies that can be used in practice: for automatic or semiautomatic data integration, for identification of mappings between semantic data structures / schemes, for identification of functional equivalence of semantically defined web services and for ontology alignment. For this purpose, applications of standards such as Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), SPARQL Protocol and RDF Query Language, Web Ontology Language (OWL) and Semantic Annotations for WSDL and XML Schema) are presented. In particular, Sesame, OpenLink Virtuoso, DBpedia and DBpedia Spotlight are presented.

The practical assumption that is followed within the course is that, in the presence of frequent data incompatibility, incompleteness and heterogeneity that takes place in real-world data integration systems, the system supporting semantic data integration should be seen as a specialized machine learning system, which, thanks to the availability of semantic data, is able to effectively reduce the workload of an expert overseeing the process of semantic data integration. In addition, the interdependence of semantic technologies and Natural Language Processing technologies (e.g. the dependency between DBpedia and DBpedia Spotlight) is considered. For this purpose selected text data processing and analysis technologies are presented, with particular emphasis on Latent Semantic Analysis (and related methods) and methods based on randomized variables indexing (especially Random Indexing and Reflective Random Indexing), semantic search technologies and technologies for integration of recommendation systems with semantic search systems.

In order to prepare the student for an objective evaluation of a semantic data integration system and to facilitate the student's own research in this field, a standard evaluation methodology that is commonly used in both academia (including the S3 and OAEI contests) and industry is presented, along with key performance measures and examples of their implementations based on the leading software libraries.

Teaching methods:

- presentation of the theory with frequent references to relevant practical examples of software implementations,
- lecture with multimedia presentation and presentations of Python programming language source code examples with their execution and rapid development/modification,
- students being asked questions during the lectures in order to provoke discussions.

Laboratory

To help the student effectively design and implement her/his own efficient and scalable semantic representation processing and integration system, examples of the use of software libraries and software packages are provided (mainly software libraries and software packages that are typically used to build machine learning systems, such as Scikit-learn). The examples include simple systems based on the RDFLib library and other Python programming language libraries, including those enabling the development of SPARQL client applications, processing of ontology data expressed in OWL, and processing Web services representations expressed in SAWSDL language.

In order to facilitate the implementation of natural language processing techniques, examples of Python scripts involving the use of NLTK and Spacy libraries are presented.

The accuracy of mappings generated by the system implemented by the student is evaluated using the data sets of the Semantic Service Selection (S3), and the Ontology Alignment Evaluation Initiative (OAEI) contests.

Teaching methods:

- Individual work on the system (mainly on its source code in Python programming language),
- Work on open source tools and software components (including those developed in Poznan University of Technology research projects) made available to students to support their work,
- At the end of each class a short presentation given by the student to present the results of her/his work followed by the lecturer?s comments and recommendations,
- At the end of the semester preparation of the report on the implementation of all the laboratory tasks.

2017 update

A major modification of the whole course description and the bibliography has been made. In particular several new topics have been introduced, including: systems for mapping between semantic data structures / schemes, matchmaking systems for semantically defined network services, and ontology alignment systems, leading systems enabling representation of semantics in accordance with Semantic Web and Linked Data concepts, software libraries and software packages supporting development of semantic data processing and integration systems, integration of semantic and NLP technologies.

## Basic bibliography:

1. John Hebeler, Matthew Fisher, Ryan Blace, Andrew Perez-Lopez, Mike Dean, Semantic Web Programming, Wiley Publishing, ISBN 978-0-470-41801-7, 2009.

2. Semantic Web, Wikipedia book, https://en.wikipedia.org/wiki/Book:Semantic_Web

3. Szwabe, A., Misiorek, P., Walkowiak, P., Multi-Relational Learning for Recommendation of Matches between Semantic Structures, in: Grana, M., Toro, C., Howlett, R.J., Jain, L.C. (Eds.), Knowledge Engineering, Machine Learning and Lattice Computing with Applications, LNCS/LNAI Volume 7828, 2013, pp. 98-107, http://link.springer.com/chapter/10.1007/978-3-642-37343-5_11, http://ncn6788.cie.put.poznan.pl/images/ncn6788_lnai.pdf

4. Szwabe A., Ciesielczyk M., Misiorek P, Blinkiewicz M., ?Application of the tensor-based recommendation engine to semantic service matchmaking,? Proceedings of The Ninth International Conference on Advances in Semantic Processing, pp. 116-125, ISBN: 978-1-61208-420-6, July 2015, Nice, France, http://www.thinkmind.org/index.php?view=article&articleid=semapro_2015_5_40_30093, http://ncn6788.cie.put.poznan.pl/images/ncn6788-semapro2015.pdf

## Additional bibliography:

1. Styperek A., Ciesielczyk M., Szwabe A., Semantic search engine with an intuitive user interface, in: Proceedings of the companion publication of the 23rd international conference on World wide web companion (WWW Companion '14). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2014, pp. 383-384. DOI=10.1145/2567948.2577203, http://dl.acm.org/citation.cfm?doid=2567948.2577203, http://ncn6788.cie.put.poznan.pl/images/ncn6788_www.pdf.

2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The elements of statistical learning : data mining, inference, and prediction, 2nd ed., Springer Series in Statistics, ISBN 978-0-387-84857-0, 978-0-387-84858-7 (e-ISBN), New York, 2016.

3. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, An Introduction to Information Retrieval, online edition, https://nlp.stanford.edu/IR-book/, https://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf, Cambridge University Press, England, 2009.

## Result of average student's workload

| Activity | Time (working hours) |
|---|---|
| 1. Paricipation in lectures. | 15 |
| 2. Participation in laboratory classes. | 30 |
| 3. Consultations. | 5 |
| 4. Preparation for laboratory classes. | 30 |
| 5. Preparation of the report. | 30 |
| 6. Preparation for the exam. | 15 |

## Student’s workload

| Source of workload | hours | ECTS |
|---|---|---|
| Total workload | 125 | 5 |
| Contact hours | 50 | 2 |
| Practical activities | 90 | 3 |