

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Reprezentowanie semantyki w sieci Web		Kod 1010335541010337157
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) (brak)	Rok / Semestr 2 / 4
Ścieżka obieralności/specjalność Technologie informatyczne	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) niestacjonarna	
Godziny Wykłady: 8 Ćwiczenia: - Laboratoria: 16 Projekty/seminaria: -		Liczba punktów 5
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (brak)		(ogólnouczelniany, z innego kierunku) (brak)
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne nauki techniczne		Podział ECTS (liczba i %) 5 100% 5 100%
Odpowiedzialny za przedmiot / wykładowca: dr inż. Andrzej Szwabe email: Andrzej.Szwabe@put.poznan.pl tel. 61 665 3958 Wydział Elektryczny ul. Piotrowo 3A 60-965 Poznań		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Ma wiedzę odpowiadającą studiom pierwszego stopnia w zakresie technologii internetowych.
2	Umiejętności:	Ma umiejętności odpowiadające studiom pierwszego stopnia w zakresie technologii internetowych.
3	Kompetencje społeczne	Ma kompetencje społeczne odpowiadające studiom pierwszego stopnia.
Cel przedmiotu: Głównym celem przedmiotu jest przedstawienie współczesnych sposobów reprezentowania semantyki w sieci Web, ze szczególnym uwzględnieniem tych technologii Semantic Web i Linked Data, które mogą zostać spożytkowane w praktyce: do automatycznej lub półautomatycznej integracji danych, identyfikacji odwzorowań między semantycznymi strukturami/schematami danych, identyfikacji funkcjonalnej równoważności semantycznie zdefiniowanych usług sieciowych i zestrajania ontologii. Przedmiot ma na celu przygotowanie studenta do samodzielnego zaprojektowania, implementacji i oceny systemu semantycznej integracji danych. Prezentowane są biblioteki programistyczne i pakiety oprogramowania które mogą być wykorzystane do sprawnego stworzenia nowoczesnego, efektywnego i skalowalnego systemu przetwarzania i integracji danych o reprezentacji semantycznej. Aby umożliwić studentowi obiektywną ocenę systemu semantycznej integracji danych i przyszłe prowadzenie własnych badań w tym zakresie, przedstawiana jest standardowa metodyka oceny wraz z kluczowymi miarami trafności odwzorowań i ich przykładowymi implementacjami przygotowanymi z użyciem wiodących bibliotek programistycznych.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		
1. ma wiedzę o aktualnych trendach dotyczących zastosowań informatyki oraz kluczowych problemów z tym związanych - [K_W06]		
2. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w zakresie informatyki - [K_W14]		
Umiejętności:		
1. potrafi - przy formułowaniu i rozwiązywaniu problemów informatycznych - integrować wiedzę z różnych dziedzin i dyscyplin naukowych - [K_U07]		
2. potrafi - pracując w zespole - sformułować specyfikację fragmentów nietypowych lub złożonych systemów informatycznych - [K_U08]		
Kompetencje społeczne:		
1. potrafi myśleć i działać w sposób kreatywny i przedsiębiorczy - [K_K01]		

Sposoby sprawdzenia efektów kształcenia
--

Wykład: egzamin pisemny (sprawdzenie wiedzy teoretycznej i umiejętności opracowania koncepcji systemu semantycznej integracji danych dla zadanego scenariusza aplikacyjnego).

Laboratorium: ocena częściowa z realizacji ćwiczenia laboratoryjnego (polegającego na zaprojektowaniu i implementacji komponentu systemu semantycznej integracji danych), ocena końcowa w połowie średnią ocen częściowych (przyznawanych na każdych zajęciach) a w połowie oceną implementacji złożonej z sukcesywnie przygotowanych komponentów i jej prezentacji w sprawozdaniu.

Treści programowe

Wykład

Przedmiot dotyczy technologii, dzięki którym zarówno ludzie, jak i systemy informatyczne są w stanie odczytywać i zapisywać zależności semantyczne między reprezentowanymi w danych obiektami ze świata rzeczywistego. Głównym celem przedmiotu jest przedstawienie współczesnych sposobów reprezentowania semantyki w sieci Web, ze szczególnym uwzględnieniem tych technologii Semantic Web i Linked Data, które mogą zostać spożytkowane w praktyce: do automatycznej lub półautomatycznej integracji danych, identyfikacji odwzorowań między semantycznymi strukturami/schematami danych, identyfikacji funkcjonalnej równoważności semantycznie zdefiniowanych usług sieciowych i zestrainowania ontologii (ang. ontology alignment). W tym celu prezentowane są zastosowania takich standardów jak Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), SPARQL Protocol and RDF Query Language (SPARQL), Web Ontology Language (OWL) i Semantic Annotations for WSDL and XML Schema (SAWSDL). W szczególności prezentowane są systemy Sesame, OpenLink Virtuoso, DBpedia i DBpedia Spotlight.

Przyjmuje się praktyczne założenie, że wobec częstego występującego w rzeczywistych systemach integracji danych zjawiska niepełności i częściowej niepoprawności danych oraz silnej heterogeniczności schematów system wspierający tę integrację jest wyspecjalizowanym systemem uczącym się (ang. Machine Learning system), który dzięki dostępności danych semantycznych jest w stanie efektywnie zredukować nakład pracy eksperta nadzorującego proces semantycznej integracji danych. Ponadto przyjmuje się celowość prezentacji technologii semantycznych wraz z pokrewnymi technologiami przetwarzania danych wyrażonych w języku naturalnym (ang. Natural Language Processing) w tym na przykładzie zależności między systemem DBpedia a systemem DBpedia Spotlight. W tym celu prezentowane są m.in. technologie eksploracji danych tekstowych - ze szczególnym uwzględnieniem metody Latent Semantic Analysis i metod pokrewnych oraz metod bazujących na indeksowaniu zrandomizowanym (w szczególności Random Indexing i Reflective Random Indexing), technologie wyszukiwania semantycznego (bazujące na przestrzeniach wektorowych) oraz zagadnienia integracji systemów rekomendacyjnych z systemami wyszukiwania semantycznego.

W celu przygotowania studenta do przeprowadzenia obiektywnej oceny systemu semantycznej integracji danych i ułatwienia prowadzenia przez studenta własnych badań w tym zakresie, przedstawiana jest standardowa metodyka oceny - powszechnie stosowana w środowisku akademickim (w tym w w.w. konkursach) i przemysłowym - wraz z kluczowymi miarami trafności odwzorowań i ich przykładowymi implementacjami przygotowanymi z użyciem wiodących bibliotek programistycznych.

Metody kształcenia

Stosowane są następujące metody kształcenia:

- prezentacja teorii z częstymi odniesieniami do praktycznych przykładów implementacji,
- wykład z prezentacjami multimedialnymi i prezentacjami przykładów kodu źródłowego w języku programowania Python, jego modyfikacji i wyników uruchomienia,
- dyskusje inicjowane w trakcie wykładu.

Laboratoria

Aby ułatwić studentowi zaprojektowanie i implementację własnego systemu przetwarzania danych semantycznych, prezentowane są przykłady zastosowania bibliotek programistycznych i pakietów oprogramowania wykorzystywanych najczęściej do budowy systemów uczących do sprawnego stworzenia nowoczesnego, efektywnego i skalowalnego systemu przetwarzania i integracji danych o reprezentacji semantycznej. W szczególności prezentowane są przykładowe implementacje obejmujące użycie biblioteki RDFLib i innych bibliotek programistycznych języka Python w tym ułatwiających budowę aplikacji klienckich SPARQL, przetwarzanie danych ontologii wyrażonych w języku OWL i przetwarzanie reprezentacji usług sieciowych (ang. web services) wyrażonych w języku SAWSDL.

W celu ułatwienia realizacji zadań obejmujących zastosowanie technik przetwarzania danych w języku naturalnym, prezentowane są przykłady kodu w języku Python stanowiące efektywne zastosowania bibliotek (NLTK) i Spacy.

Ocena trafności odwzorowań generowanych przez implementowany przez studenta system dokonywana jest z użyciem zbiorów danych stosowanych z konkursach Semantic Service Selection (Annual International Contest S3 on Semantic Service Selection, Retrieval Performance Evaluation of Matchmakers for Semantic Web Services), Ontology Alignment Evaluation Initiative (OAEI).

Metody kształcenia

Stosowane są następujące metody kształcenia:

- indywidualna praca studenta wspierana przygotowaniem przed zajęciami (m.in. zapoznaniem się z udostępnionymi przykładami implementacji),
- praca z narzędziami o otwartym kodzie źródłowym (ang. open source) i składnikami oprogramowania opracowanymi w projektach badawczych Politechniki Poznańskiej udostępnianymi studentom przez prowadzącego zajęcia,
- na koniec każdego zajęć krótkie przedstawienie przez studenta uzyskanych wyników (każdorazowo uzupełnione odpowiedziami wykładowcy i jawną oceną wyników dokonaną przez wykładowcę),
- na koniec semestru przygotowanie sprawozdania z realizacji wszystkich zadań.

Aktualizacja w 2017 r.

Dokonano poważnej modyfikacji całego opisu kursu i bibliografii. Uszczegółowiono opis zagadnień dotąd prezentowanych i wprowadzono kilka nowych - takich jak: systemy identyfikacji odwzorowań między semantycznymi strukturami/schematami danych, systemy identyfikacji funkcjonalnej równoważności semantycznie zdefiniowanych usług sieciowych i systemy zestrainowania ontologii. Dodano prezentację wiodących systemów umożliwiających reprezentowanie semantyki zgodnie z koncepcjami Semantic Web i Linked Data oraz bibliotek programistycznych umożliwiających implementację systemów semantycznej integracji danych

Literatura podstawowa:

1. John Hebel, 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>

Literatura uzupełniająca:

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.

Bilans nakładu pracy przeciętnego studenta

Czynność	Czas (godz.)
1. Wykłady	15
2. Laboratoria	30
3. konsultacje	5
4. Przygotowanie do zajęć laboratoryjnych	30
5. Przygotowanie sprawozdania	30
6. Przygotowanie do egzaminu	15

Obciążenie pracą studenta

forma aktywności	godzin	ECTS
Łączny nakład pracy	125	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	50	2
Zajęcia o charakterze praktycznym	90	3