

Wykrywanie wiedzy w dużych zbiorach danych: przykład personalizacji inżynierii ontologicznej

Cezary Chudzian, Janusz Granat,
Edward Klimasara, Jarosław Sobieszek,
Andrzej P. Wierzbicki

W artykule, po przedyskutowaniu szeroko rozumianego pojęcia inżynierii wiedzy, a w szczególności inżynierii ontologicznej, koncentrującej się na wykrywaniu wiedzy w dużych zbiorach tekstu, omówiono rezultaty prac wykonanych w Instytucie Łączności PIB w ramach projektu badawczego zamawianego „Usługi i sieci teleinformatyczne następnej generacji – aspekty techniczne, aplikacyjne i rynkowe”. Prace wstępne i rozpoznawcze doprowadziły do sformułowania założenia na przyjazny, spersonalizowany system wspomagania wyszukiwania wiedzy w dużych zbiorach tekstu. Efektem końcowym zaś jest pilotowy system PrOnto wykorzystujący narzędzia inżynierii ontologicznej w dwóch aspektach: jednym – wspomagającym wyszukiwanie informacji w dokumentach elektronicznych na podstawie osobistej ontologii użytkownika i drugim – wspomagającym współdzielenie źródeł wiedzy przy wykorzystaniu modeli ontologicznych. Artykuł kończy wnioski co do stosowalności różnych metod wykorzystania narzędzi inżynierii ontologicznej oraz sposobów kontynuacji prac.

inżynieria wiedzy, inżynieria ontologiczna, personalizacja ontologii, human centered computing

Wprowadzenie

W ciągu ostatniej dekady lat, szczególnego znaczenia w telekomunikacji oraz szerzej – w wykorzystaniu internetu, nabrało wynajdywanie formuł logicznych i relacji, dotyczących wybranych informacji oraz modeli wiedzy zawartej w bardzo dużych zbiorach danych, bądź zgromadzonych przez operatorów telekomunikacyjnych, udostępnianych urzędом regulacyjnym, bądź po prostu dostępnych w internecie.

Prace w tym zakresie są prowadzone w Instytucie Łączności już ponad dziesięć lat. Rozwijanych jest kilka głównych działów technik inżynierii wiedzy^①.

- Wykorzystanie zaawansowanych narzędzi i metod
 - logiki matematycznej,
 - inżynierii ontologicznej,
 - optymalizacji i podziału zbiorów danych do rozpoznawania wzorców,
 - wykrywania zdarzeń i anomalii,
 - wielokryterialnej teorii decyzji.
- Wykorzystanie różnorodnych zaawansowanych metod statystycznych.

Wszystkie one mogą służyć wykrywaniu wiedzy w dużych zbiorach danych [1].

^① Terminy angielskie to „data mining” lub „knowledge mining”, związane z tym inne pojęcia to „knowledge management”, „knowledge engineering” i „knowledge science”. „Naukoznawstwo” ma inne (filozoficzne) tradycyjne znaczenie w języku polskim, „zarządzanie wiedzą” zaś jest dzisiaj głównie przedmiotem nauk o zarządzaniu, dlatego postanowiono wykorzystywać w tym zakresie pojęcie inżynierii wiedzy w szerokim znaczeniu tego słowa, nie ograniczając go, jak to bywa w tradycji akademickiej, do metod sztucznej inteligencji i algorytmów uczenia się maszyn.

Tak szeroko rozumiana inżynieria wiedzy może być wykorzystywana w bardzo różnorodnych dziedzinach, np. w opracowaniach dla operatorów telekomunikacyjnych, a zaawansowane metody statystyczne mogą być przydatne do analizy wartości różnych wskaźników stanu rozwoju społeczeństwa informacyjnego w Polsce czy na Mazowszu. W tym artykule omówiono głównie metody inżynierii ontologicznej we wspomaganiu wykrywania wiedzy i zarządzania wiedzą.

Trzeba tu dodać jeszcze jedno wyjaśnienie. Metody inżynierii ontologicznej koncentrują się zazwyczaj, jak to wynika z dotychczasowych prac nad sztuczną inteligencją, na automatyzacji wydobywania wiedzy z dużych zbiorów tekstu, preferencje użytkownika zaś mogą być uwzględniane, ale zazwyczaj w niewielkim zakresie. Specyfika prac referowanych w tym artykule jest jednak odmienna. Zakładając nadrzędną pozycję użytkownika – co uzasadniono bardziej szczegółowo dalej – koncentrowano się na *radikalnej personalizacji* osobistej ontologii użytkownika, polegającej na łączeniu intuicyjnie określonej przez użytkownika *ontologii odgórnej z ontologią oddolną* wynikającą z automatycznej analizy dużych zbiorów tekstu.

Inżynieria wiedzy a inżynieria ontologiczna

W związku z narastającą ilością informacji i danych cyfrowych gromadzonych bądź to w internecie, bądź u operatorów telekomunikacyjnych i w innych instytucjach czy przedsiębiorstwach oraz w uniwersytetach i instytutach badawczych, ważnym problemem stała się kwestia wyszukiwania w wielkich zbiorach danych nie tylko informacji interesujących, ale więcej – przydatnych relacji między takimi informacjami, czyli inaczej mówiąc – *wiedzy ukrytej w wielkich zbiorach danych*.

Używa się tu celowo określenia *wiedzy ukrytej (tacit knowledge)*, choć stosuje się ono w zasadzie do wiedzy trudnej do wyrażenia słowami, przedślownej, ukrytej zazwyczaj w umyśle ludzkim^① [5-8]. Ale rzecz w tym, że wiedza przedślowna, jeszcze nie wyrażona słowami czy wzorami, jest zawarta także w wielkich zbiorach danych, a trudnym zadaniem jest właśnie wydobyć jej z tych zbiorów, czyli dokonanie tego, co w nomenklaturze anglojęzycznej nazywane jest *data mining* czy też raczej *knowledge mining*, i wyrażenie tej wiedzy ukrytej słowami, wzorem, regułą logiczną czy inną formą modelu.

W niektórych pracach źródłowych używa się wręcz pojęcia *knowledge science*; jednakże – jak wspomniano – w języku polskim *naukoznawstwo* ma inne tradycyjne znaczenie, o odcieniu filozoficznym obejmującym epistemologię i inne dziedziny pokrewne. Dziedziną pokrewną zarówno do naukoznawstwa jak i do wydobywania wiedzy z dużych zbiorów danych jest *zarządzanie wiedzą*, ale ono, choć historycznie wyrosło z informatyki (zob. np. [9]), jest traktowane dzisiaj raczej jako przedmiot nauk o zarządzaniu. Z tych wszystkich względów używa się raczej pojęcia *inżynieria wiedzy* – w stosunku do konstrukcji i wykorzystania informatycznych narzędzi wydobywania wiedzy ukrytej z dużych zbiorów danych, lub nawet szerzej, obejmując tym pojęciem także informatyczne narzędzia przetwarzania dużych tekstów, zatem informacji słownej. Takie znaczenie pojęcia inżynierii wiedzy jest znacznie szersze, niż jego tradycyjne znaczenie akademickie, traktujące inżynierię wiedzy wąsko, jako związaną ze sztuczną inteligencją i metodami automatycznego uczenia się maszyn – przy czym to szersze rozumienie oczywiście obejmuje to rozumienie węższe, ale kładzie większy nacisk na rolę człowieka i użytkownika systemów informatycznych w procesach wydobywania czy przetwarzania wiedzy.

^① Zazwyczaj, choć nie tylko, gdyż do wiedzy ukrytej można zaliczyć intuicyjne dziedzictwo ludzkości, obejmujące m.in. sądy syntetyczne a priori [2] czy horyzonty hermeneutyczne (zob. np. [3]), w istocie wyrażające przekonania intuicyjne utrwalane przez systemy edukacyjne, a także emocjonalne dziedzictwo ludzkości, obejmujące m.in. nieświadomość zbiorową ([4]) wraz z jej częściami – archetypami i mitami ludzkości, czy też ładunek emocjonalny wszystkich filmów – zatem wiedza ukryta może być zawarta nie tylko w umyśle człowieka (zob. [5]).

Przy tym szerszym jej rozumieniu, inżynierię wiedzy można podzielić na kilka podstawowych działów.

Dział I Wąsko rozumiana inżynieria sztucznej inteligencji i automatycznego uczenia się.

Dział II Inżynieria wydobywania wiedzy ukrytej z dużych zbiorów danych.

Dział III Inżynieria przetwarzania tekstu, czyli także wydobywania wiedzy, ale wyrażonej w formie słownej.

Dział I nie będzie tu omawiany, gdyż poświęcone są mu obszerne monografie, (np. [10]). Dział II ma na celu wyrażanie wiedzy ukrytej zawartej w dużych zbiorach danych w formie użytecznych dla użytkownika modeli: logicznych, statystycznych, decyzyjnych. Opiera się zatem na takich dziedzinach podstawowych, jak logika, statystyka, wielokryterialna teoria decyzji, itp. Jest on oczywiście związany z Działem I, wykorzystując do pewnego stopnia automatyczne uczenie się, ale przy założeniu znacznie większej roli człowieka w wydobywaniu i przetwarzaniu wiedzy, tworzeniu modeli wiedzy. W wielu zastosowaniach, istotna jest także interpretacja słowna takich modeli wiedzy, zgodna z wymaganiami użytkownika, mającymi zazwyczaj także charakter wiedzy ukrytej, którą dopiero trzeba przekształcić w słowa.

Dział III ma na celu wynajdywanie czy też wybór wiedzy tekstowej, jawnej, istotnej dla użytkownika a zawartej w dużych zbiorach tekstowych. Dziedziny podstawowe dla tego działu to *inżynieria ontologiczna* (konstruowanie taksonomii uzupełnionych o różnorodne relacje logiczne w dużych zbiorach tekstu), *sieci semantyczne (semantic web)*, inżynieria wyszukiwarek internetowych, itp. W zastosowaniach, istotna okazuje się interpretacja tej wybranej wiedzy tekstowej przez użytkownika, zatem znów zgodnie z jego wiedzą ukrytą, tak jak w kręgu hermeneutycznym (zob. [11]) czy w hermeneutycznej spirali kreowania wiedzy (zob. [12]).

Inżynieria wiedzy może być pomocna w zarządzaniu wiedzą. Jest oczywiste, że samo pojęcie zarządzania wiedzą może być interpretowane – zwłaszcza przy uwzględnieniu dużego znaczenia wiedzy ukrytej w umysłach ludzi – jako absurdalne czy wewnętrznie sprzeczne. Jest jednak faktem historycznym, że pojęcie *knowledge management* zostało użyte najpierw w odniesieniu do narzędzi inżynierii wiedzy mających zapewnić ciągłość prac nad oprogramowaniem przez firmę DEC (*Digital Equipment Corporation*) we wczesnych latach osiemdziesiątych (choć podobne narzędzia, bez użycia pojęcia *knowledge management*, były już wcześniej stosowane przez firmę IBM). Dopiero w latach dziewięćdziesiątych pojęcie to zaczęło się pojawiać w naukach o zarządzaniu, co doprowadziło, z jednej strony, do ogromnej kariery tego niezbyt precyzyjnego hasła, z drugiej strony jednak do licznych kontrowersji (zob. np. [9]). Stąd też tradycja traktowania zarządzania wiedzą jako części inżynierii software'owej jest szacowna. Ale nauki o zarządzaniu, które wdrożyły to pojęcie o dekadę później, nadały mu jednak znacznie szersze znaczenie i opublikowały ogromną literaturę z tym związaną. Tak więc istnieją dziś dwa przeciwne poglądy, jak interpretować to pojęcie [13], [14]:

- Jako *zarządzanie informacją istotną dla działań związanych z wiedzą*, z naciskiem na systemy informacyjne, bazy i składnice danych, wydobywanie wiedzy z danych, groupware, systemy obiegu dokumentów, analizę tekstów i inżynierię ontologiczną, etc.
- Jako *zarządzanie procesami związanymi z wiedzą*, z naciskiem na teorię organizacji, uczenia się, rodzaje wiedzy oraz procesów kreowania wiedzy.

Pierwszy pogląd jest naturalnie reprezentowany przez informatyków i nauki czy techniki pokrewne, natomiast drugi przez socjologów, teoretyków zarządzania, filozofów czy psychologów i dominuje w naukach o zarządzaniu. Reprezentanci drugiego poglądu zarzucają zwolnikom pierwszego trakto-

wanie wiedzy jako obiektu podczas gdy trzeba ją postrzegać jako wiedzę związaną z procesami, i twierdzą, że zarządzanie wiedzą to zarządzanie ludźmi. Wszystko to prawda (zob. także [15], [16]), ale oba te poglądy są jednostronne. Chociaż trzeba przyznać, że zarządzanie wiedzą nie może być zredukowane do zarządzania informacją, to jednak bardzo często słuszne poglądy powodują nadmierne uproszczenia i przeoczenia. Problem jest bardziej złożony, ponieważ zarządzanie wiedzą rozpoczęło się od technik informacyjnych i nie może być bez nich kontynuowane; a istota problemu polega na tym, że zarządzanie ludźmi musi być dziś rozumiane jako zarządzanie pracownikami wiedzy, a ci bardzo często reprezentują obecnie techniki informacyjne.

Co więcej, zarządzanie wiedzą w instytucjach komercyjnych – zwykle wielkich korporacjach, gdyż małe firmy dziś jeszcze bardzo rzadko stać na kosztowne systemy zarządzania wiedzą – ma zupełnie odmienny charakter, niż zarządzanie wiedzą w instytucjach badawczych czy akademickich^①, (zob. [5], [8]). Zarządzanie wiedzą w instytucji administracyjnej czy regulacyjnej może mieć większe podobieństwo do zarządzania wiedzą w instytucji badawczej, niż w instytucji komercyjnej – i nawet wielkie instytucje regulacyjne będą szukały jak najtańszych rozwiązań zarządzania wiedzą. Stąd całe prace grupy tematycznej zajmującej się zagadnieniem wykrywania wiedzy opierały się na założeniu maksymalnego wykorzystania oprogramowania swobodnego, darmowego (*free software*) oraz na specyfice lokalnego zarządzania wiedzą.

Działalność naukowo-badawcza charakteryzuje się szczególnymi wymaganiami w zakresie zarządzania wiedzą. W odróżnieniu od potrzeb aktywności produkcyjnej, czy usługowej typowo zorientowanych na wspomaganie działań operacyjnych, jednym z najistotniejszych zadań działalności naukowej jest eksploracja potencjalnych kierunków rozwoju. Tradycyjne systemy zarządzania wiedzą, powstałe w dużej mierze w odpowiedzi na rzeczywiste potrzeby przedsiębiorstw, skupiły się w przeważającym stopniu na wspomaganie praktycznych aspektów ich działalności, pomijając właściwie zastosowania charakterystyczne dla działalności naukowo-badawczej.

Procesy opisujące powstawanie wiedzy w organizacjach (w tym także naukowo-badawczych) zostały zebrane w ramach teorii „kreatywnego środowiska” [8], będącej syntezą prac zapoczątkowanych na gruncie operacyjnym [17]. Jednym z elementów tej teorii jest model działalności naukowej nazwany *potrójną helisą* opisujący typowy sposób powstawania wiedzy w instytucjach naukowo-badawczych. Składa się on z trzech współwystępujących procesów, tzw. spiral tworzenia wiedzy, przedstawiających:

- badania literaturowe (spirała EAIR),
- eksperymenty (spirała EEIS),
- dyskusje (spirała EDIS).

Nazwy spiral są akronimami pochodzącymi od angielski nazw czynności składowych danego procesu. Cykliczna natura spiral oddaje ciągłość i powtarzalność procesu tworzenia wiedzy.

Dalej zasadniczym tematem będzie wspomaganie procesu badań literaturowych opisanego przez spiralę EAIR, zwaną też z racji swoich podstaw filozoficznych [11] *spiralą hermeneutyczną*. Składa się ona z czterech etapów:

^① W instytucjach akademickich główne motywacje są związane z indywidualnymi celami twórcy wiedzy (powiększanie reputacji naukowej, zdobywanie stopni i tytułów, w mniejszym stopniu cele komercyjne) oraz z powiększaniem intelektualnego dziedzictwa ludzkości, zatem z indywidualną oraz publiczną własnością wiedzy. W instytucjach komercyjnych główne motywacje związane są z celami grupowymi (dochód i reputacja przedsiębiorstwa), zatem z korporacyjną prywatyzacją wiedzy. Dlatego też uproszczony sąd postmodernistycznej socjologii o tym, że skoro uczelnie są producentami wiedzy, zatem powinny być zarządzane tak jak przedsiębiorstwa komercyjne, jest dalece chybiony.

- *Enlightenment* – ten etap zaczyna się od pomysłu, który jest uważany za warty dalszej eksploracji i obejmuje proces wyszukiwania i gromadzenia potencjalnych źródeł informacji,
- *Analysis* – to etap racjonalnej analizy materiałów, które zostały uznane za istotne,
- *Hermeneutic immersion* – w tym etapie pomysły zanalizowane w poprzednim etapie zaczynają być pojmowane intuicyjnie,
- *Reflection* – to etap intuicyjnego rozważania nowych pomysłów.

Ze spiralą EAIR związane jest też pojęcie *agenta hermeneutycznego* [8], czyli narzędzia wspierającego jej kolejne etapy, mającego służyć badaczom w znajdowaniu i analizie źródeł potrzebnej im informacji. Racjonalna natura dwóch pierwszych faz spirali hermeneutycznej powoduje, że są one dużo bardziej podatne na próby ich wspomaganie przy użyciu narzędzi komputerowych. Zwłaszcza pierwsza z nich jest coraz bardziej istotna w obliczu szybko rosnącej liczby potencjalnych źródeł informacji, co jest związane w dużym stopniu z rozwojem internetu. Warto zwrócić uwagę, że termin „przeciążenie informacją” został użyty po raz pierwszy w 1970 roku [18], a więc jeszcze przed pierwszymi pracami opisującymi protokoły TCP/IP, na długo przed rozpowszechnieniem internetu. Obecnie liczba i zawartość informacyjna źródeł wiedzy dostępnych przez internet przewyższa inne dostępne kanały.

Analiza tekstu i inżynieria ontologiczna rozwijały się równolegle do *zarządzenia wiedzą*, chociaż niewątpliwie mogą być one traktowane jako ważne narzędzia tej obszerniejszej dziedziny (zob. np. [19]). Słowo *ontologia* było zapożyczone z filozofii, gdzie oznacza *teorię bytów* lub systematyczną *analizę bycia* (zob. np. [20]); w informatyce nadano mu inne znaczenie, odpowiadające raczej klasyfikacji bytów oraz słów je reprezentujących. Obecnie, w technikach informacyjnych traktuje się ontologię jako wzbogaconą taksonomię, słownik z hierarchią i innymi wzajemnymi relacjami pojęć, ewentualnie dalej wzbogacony o sformalizowane relacje logiczne między pojęciami lub ich klasami. W ciągu lat dziewięćdziesiątych nastąpiła znaczna ewolucja inżynierii ontologicznej, związana z pojęciem *semantic web*, oparta na założeniu^①, że współczesna sieć powinna zawierać zbiór wiedzy odpowiadający dorobkowi intelektualnemu ludzkości, a zatem współczesne techniki informacyjne powinny z niej móc wydobyć całą ludzką wiedzę w postaci uniwersalnej ontologii. Traktowane jako środki reprezentacji i wspólnego rozumienia wiedzy o różnych dziedzinach świata, ontologie odgrywają dzisiaj ważną rolę w wielu zastosowaniach, takich jak rozwój systemów informacyjnych, organizacja zawartości stron internetowych, kategoryzacja produktów komercyjnych, słowniki standardowe w określonych dziedzinach, itp. (zob. np. [21]-[24]).

Także w tej dziedzinie można zaobserwować różne kontrowersje, związane z kilkoma przeciwstawnymi podejściami do konstrukcji, wykorzystania i interpretacji ontologii. Jest wiele metod konstrukcji ontologii, (zob. np. [25]), można wśród nich wyróżnić metody konstrukcji *ontologii lekkich* czy też *o lekkiej strukturze* (*lightweight ontologies*, zawierających prostą, np. hierarchiczną, strukturę pojęć) oraz *ontologii ciężkich* czy też *o ciężkiej strukturze* (*heavyweight ontologies*, obejmujących złożone relacje pomiędzy pojęciami, ich logiczną strukturę oraz możliwości formalizacji rozumowania); konstrukcję *ontologii lokalnych* charakteryzujących pojęcia używane przez daną grupę, np. badaczy czy w danej sferze kulturowej (różne sfery kulturowe używają często tych samych słów, jak *ontologia*, w absolutnie odmiennych znaczeniach) w przeciwstawieniu do *ontologii uniwersalnych* próbujących reprezentować wszystkie materiały danej dziedziny bądź wszystkie publikowane w internecie. Można też mówić o konstrukcji ontologii *od początku* (*from scratch*), *przez adaptację* (*reuse*) lub *automatycz-*

^① Założenie to jest dyskusyjne, (zob. np. [5]) o roli komponentów intuicyjnych i emocjonalnych w dziedzictwie intelektualnym ludzkości, dyskutowanej też w jednym z poprzednich odnośników, oraz dalsze uwagi w tym artykule o powodach radykalnej personalizacji ontologii osobistych.

nie (stosując automatyczne metody inżynierii ontologicznej, (zob. [24]); podział taki nie jest jednak w pełni precyzyjny, gdyż dobre narzędzia inżynierii ontologicznej są zawsze *półautomatyczne*, zakładają pewną interakcję z użytkownikiem (konstruktorem ontologii), przy czym zasadniczym problemem jest charakter i sposób takiej interakcji. Wśród najbardziej zaawansowanych są prace *Standard Upper Ontology Working Group* (SUO WG), skierowane na "forming an upper ontology whose domain is all of human consensus reality" wraz ze związaną z nimi ontologią CYC (zob. np. [26]); jest to niewątpliwie ciekawa próba budowy uniwersalnego słownika, można jednak mieć do niej wiele wątpliwości z perspektywy zastosowań lokalnych, gdzie popularność użycia danego pojęcia w internecie nie musi decydować o znaczeniu i rozumieniu tego pojęcia.

Związane jest to jeszcze z jednym, zasadniczym rozróżnieniem sposobów konstrukcji ontologii: *odgórną konstrukcją ontologii* (*top down*), wynikającą z doświadczenia ekspertów, ich intuicji i emocjonalnych przekonań, oraz *oddolną konstrukcją ontologii* (automatyczną lub półautomatyczną, wynikającą z jawnie zapisanych tekstów). Lokalna konstrukcja ontologii była zazwyczaj odgórna, ale obecny stan rozwoju inżynierii ontologicznej umożliwia kombinację metod odgórnych i oddolnych, problemem jest tylko sposób ich połączenia. Można na to spojrzeć z perspektywy prób łączenia *wiedzy ukrytej* (*tacit knowledge*, wiedzy intuicyjnej i emocjonalnej) z wiedzą jawną (*explicit knowledge*, wiedzy racjonalnej), które według [5], [7] są podstawą wszelkich metod kreowania wiedzy. Z tej perspektywy, uniwersalne ontologie są mało wartościowe do zastosowań lokalnych, gdyż nie uwzględniają wiedzy ukrytej lokalnej grupy ekspertów.

Perspektywę taką potwierdza techniczna i ewolucyjna teoria intuicji [27], [28], która wykorzystuje współczesną wiedzę telekomunikacji i informatyki dla wykazania, że użycie języka (zatem także logiki) uprościło co najmniej dziesięć tysięcy razy^① postrzeganie i rozumowanie obrazowe oraz ogólniej immanentne (wszystkimi zmysłami). W wyniku tego pozostał ogromny „nadmiar” mózgu^②, wykorzystywany w rozmaitych funkcjach życiowych, umiejętnościach ukrytych, także niekiedy we wnioskowaniu intuicyjnym, egzystencjalnym, transcendentnym, etc. Nadmiar ten wyraża się oszacowaniem, że tylko nie więcej niż 0,1 promila neuronów mózgu człowieka zaangażowana jest w myślenie logiczne i argumentację słowną. Dlatego ludzka intuicja może być znacznie potężniejsza od argumentacji logicznej. Wynika stąd jednak także potrzeba radykalnej personalizacji osobistych ontologii użytkowników narzędzi ontologicznych, polegającej na zwiększeniu roli ich intuicji w definiowaniu ich ontologii, rozpoczynaniu od lokalnej ontologii odgórnej.

Z tej perspektywy właśnie warte zainteresowania są nowe podejścia do tworzenia ontologii (od początku lub przez adaptację, o lekkiej strukturze) dla lokalnej grupy badaczy, łączącymi podejścia odgórne i oddolne, półautomatyczne.

W grupie tematycznej zajmującej się wykrywaniem wiedzy związanej z projektem PBZ „Usługi i sieci teleinformatyczne następnej generacji – aspekty techniczne, aplikacyjne i rynkowe” zakładano początkowo szersze wykorzystanie i rozwój metod inżynierii wiedzy dla wspomaganie decyzji regulacyjnych, lecz cel ten musiał ulec zawężeniu w wyniku zasadniczego ograniczenia przyznanego finansowania tego tematu. Dlatego wybrano dziedzinę częściową, która jednak rokuje szybkie zastosowania w administracji podejmującej decyzje regulacyjne, a mianowicie wykrywanie wiedzy w dużych

^① Pasma niezbędne do przekazywania obrazów jest co najmniej 100 razy większe niż pasmo niezbędne dla przekazywania mowy, a złożoność obliczeniowa przetwarzania takich dużych zbiorów informacji jest nieliniowa, przyjęcie zależności kwadratowej jest tu oszacowaniem łagodnym. Stąd obraz jest wart co najmniej dziesięć tysięcy słów. Zatem w ewolucyjnym rozwoju człowieka, po etapie mowy, nastąpił ogromny skrót ewolucyjny.

^② Niektórzy filozofowie mówili o nadmiarze umysłu nad mózgiem, ale zgodnie z omawianą tu ewolucyjną teorią intuicji jest to właśnie nadmiarowość mózgu, czasami wykorzystywana jako nadmiar umysłu.

zbiorach danych tekstowych oraz zarządzanie wiedzą w określonej instytucji, nawiązując do Działu III w przedstawionej klasyfikacji metod inżynierii wiedzy, czyli analizę tekstu oraz inżynierię ontologiczną. Z tego samego powodu, zdecydowano się ograniczyć wstępne zastosowania i testy do instytucji badawczej, tj. Instytutu Łączności.

Rezultaty prac grupy tematycznej zajmującej się zagadnieniem wykrywania wiedzy

Prace wstępne i konstrukcja przykładowych ontologii

We współpracy z IIASA (*International Institute for Applied Systems Analysis*) oraz JAIST (*Japan Advanced Institute for Science and Technology, School of Knowledge Science*) przeprowadzono bardzo obszerny przegląd literatury. Z przeglądu tego jednak wynika, że chociaż kwestia łączenia metod oddolnych i oddolnych w tworzeniu ontologii była podjęta (np. [29]), nie zaproponowano jednak specjalnych metod takiego połączenia. Wyjątkiem jest podejście użyte wspólnie ze współpracownikami japońskimi, a relacjonowane m.in. w [16]. W podejściu tym wyrażenie wiedzy ukrytej ekspertów wspomaganie jest *refleksją hermeneutyczną*, to jest refleksją tych ekspertów nad świadomie poszerzoną dziedziną obejmującą opracowywaną ontologię. Inne podejście, też mające na celu wyrażenie wiedzy ukrytej oraz też prezentowane w [16], wynikające z doświadczeń prac prowadzonych w Instytucie Łączności, to *refleksja organizacyjna*, czyli refleksja ekspertów nad strukturą organizacyjną Instytutu i prowadzonych w nim badań. Rozważano też inne metody eksplikacji wiedzy ukrytej ekspertów, takie jak tworzenie *map umysłu (mind mapping)*, ale te są w mniejszym stopniu kontrolowane bezpośrednio przez eksperta, podczas gdy podstawowym założeniem eksplikacji wiedzy ukrytej powinna być *suwerenna rola eksperta* wobec komputera.

Dokonano porównania kilku narzędzi inżynierii ontologicznej, pozyskanych z sieci lub dostosowanych przez prace własne, w zastosowaniu do zbiorów danych obejmujących bądź anglojęzyczne publikacje w *Journal of Telecommunications and Information Technology (JTIT)*, bądź teksty polskojęzyczne związane z działalnością Instytutu Łączności. O ile struktura ontologii wykryta w tych pierwszych tekstach wskazywała na tylko bardzo ogólne relacje (np. takie, że sieć składa się z węzłów i łączy), o tyle w drugich tekstach miała ona już pewne podobieństwo ze strukturą organizacyjną IŁ, (zob. [30], [31]). Tak więc ze wstępnego porównania różnych metod wykrywania oraz łączenia wiedzy ukrytej z jawną przy tworzeniu ontologii wynika, że rozszerzona metoda refleksji organizacyjnej – zaczynająca się od refleksji nad strukturą organizacyjną, ujmująca też przyjętą wizję i strategię instytucji, w razie potrzeby pogłębiona o refleksję hermeneutyczną (zob. [5], [8]) – jest najbardziej skuteczna w przygotowaniu ekspertów tworzących ontologię lokalnej instytucji do interakcji i korekt wyników pracy półautomatycznych narzędzi oddolnej inżynierii ontologicznej.

Przykładowe ontologie instytucji badawczej przedstawione zostały w [16]. Trzeba jednak podkreślić, że algorytmy wykrywania wiedzy dla potrzeb instytucji badawczej obejmują nie tylko konstrukcję ontologii, lecz także systemy wspomaganie zarządzania wiedzą wykorzystujące narzędzia inżynierii ontologicznej.

Narzędzia zarządzania wiedzą

Opracowano kilka narzędzi, mających charakter rozwiązań wstępnych. Obejmowały one: system katalogowania obiektów, system zarządzania dokumentami i procesami, oraz system OntoGen PL.

Główną funkcją systemu katalogowania obiektów było tworzenie katalogu obiektów i zapewnienie możliwości łatwego wyszukiwania w nim potrzebnych danych. Z jednej strony można spojrzeć na tę funkcjonalność jak na uproszczony system zarządzania wiedzą, z drugiej jak na kooperacyjne podejście do idei agenta hermeneutycznego, (zob. [8]). Wyszukiwanie danych oparte jest na współpracy użytkowników systemu przy opisie informacji.

Katalogowanymi obiektami były pojęcia związane z działalnością publikacyjną instytucji badawczych, tzn. osoba, artykuł, instytucja i czasopismo, połączone typowymi relacjami, takimi jak autorstwo, czy zatrudnienie. Zaproponowane rozwiązanie jest jednak dużo bardziej uniwersalne. Umożliwia ono właściwie dowolną zmianę dziedziny problemu, jeżeli tylko jego opis da się wyrazić w przyjętym meta-modelu obiektowym. Metamodel składa się z klas opisywanych atrybutami i powiązanych relacjami „wiele do jednego” i „wiele do wielu” i, mimo swej prostoty, jest na tyle ogólny, że można w nim opisać wiele rzeczywistych systemów.

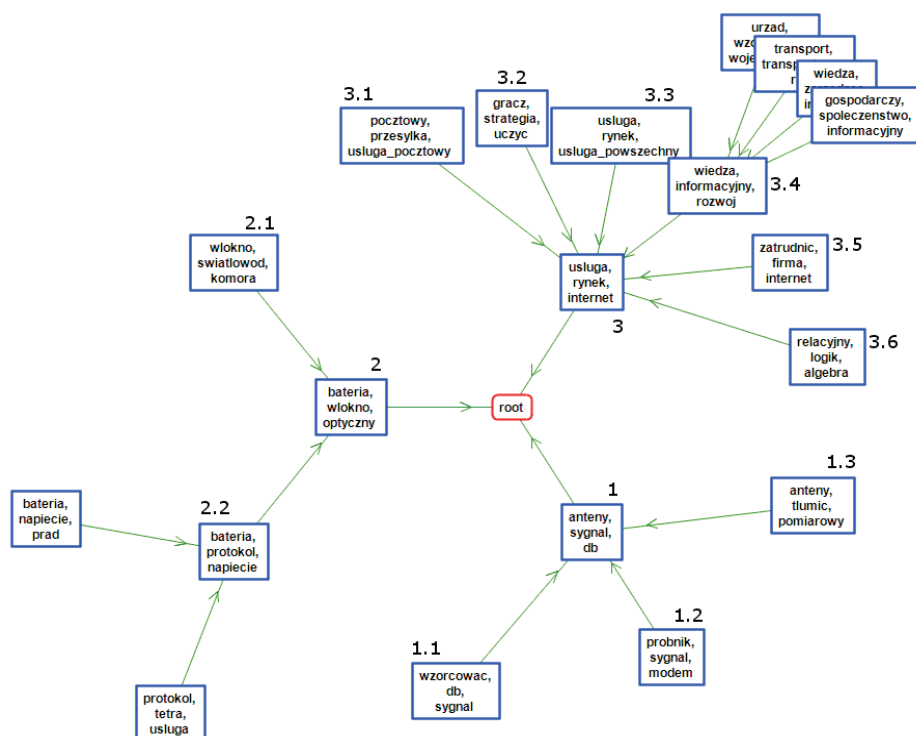
Przyjęte podejście jest przykładem tworzenia oprogramowania kierowanego modelem. Podstawowym elementem jest tu definicja modelu problemu oparta na stworzonym prostym języku opisu systemów katalogowych. Na tej podstawie automatycznie jest generowany kod programu i interfejs użytkownika.

System zarządzania dokumentami i procesami (SZDP) można traktować jak prototyp systemu zarządzania wiedzą do wspomagania kreatywności i wymiany wiedzy w środowiskach badawczych i naukowych. Głównym zadaniem SZDP jest przechowywanie i udostępnianie dokumentów w kontekście zachodzących w instytucji procesów i osób je realizujących. Dokumenty są prezentowane na tle projektów badawczych i struktury organizacyjnej instytucji.

Na poziomie metadanych, typy lub klasy procesów są definiowane na podstawie szablonów opisujących typy zadań tworzących proces, z przypisanymi wzorcami dokumentów potwierdzających ich realizację i domyślnymi terminami wykonania. Tworzenie procesu wymaga określenia jego typu i precyzyjnego przypisania terminu realizacji zadań wchodzących w jego skład oraz wyznaczenia osób lub grup osób odpowiedzialnych za zadania. Wykonanie zadania jest każdorazowo potwierdzane dostarczeniem właściwego dokumentu.

W ten sposób tworzone jest repozytorium dokumentów oraz jest gromadzona wiedza na temat realizowanych projektów badawczych i pracowników, wraz z wzajemnymi powiązaniem między wszystkimi opisywanymi obiektami, co stanowić może podstawę do analizy działalności pracowników, ich doświadczenia projektowego, dorobku publikacyjnego, czy też do tworzenia bazy dobrych praktyk projektowych.

System *OntoGen PL* to modyfikacja systemu *OntoGen* taka, aby obsługiwał on dokumenty w języku polskim z uwzględnieniem specyficznej fleksji polskiej. *OntoGen* (<http://ontogen.ijs.si/>) jest bezpłatnym programem służącym do budowy ontologii dziedziny na podstawie związanego z nią zbioru dokumentów. Ontologia powstaje na drodze realizacji interaktywnego procesu wydobywania pojęć i relacji przy wykorzystaniu metod eksploracji danych występujących w postaci tekstowej (*text mining*). Więcej na temat samego narzędzia i wyników uzyskanych przy jego zastosowaniu można znaleźć w pracy [32]. Obok konstrukcji ontologii odgórnej (w formie hierarchicznej) podjęto też próbę konstrukcji oddolnej ontologii dla Instytutu Łączności, na podstawie zbioru artykułów zawartych w wydawanym przez Instytut czasopiśmie *Journal of Telecommunications and Information Technology*. Analiza wykazała jednak, że teksty te nie są w pełni reprezentatywne dla profilu placówki, a w wykorzystaniu wewnętrznych raportów z prac prowadzonych w Instytucie, nie publikowanych w czasopiśmie, stanął na przeszkodzie brak wsparcia *OntoGen* dla języka polskiego, w którym tworzone są opracowania. Nota bene, niewiele jest też narzędzi komercyjnych, które sprawdzają się przy komplikacjach, jakie wprowadza język polski.



Rys. 1. Przykładowa ontologia oddolna zagadnień badawczych IŁ

W celu weryfikacji możliwości poprawy rezultatów osiągniętych w [32], przez uwzględnienie dokumentów polskojęzycznych, utworzone zostało środowisko programowe o roboczej nazwie OntoGen PL. Podstawową jego cechą jest wykorzystanie w całości bezpłatnego oprogramowania. Część zadań związana z przetwarzaniem danych jest realizowana przez autorski kod napisany w języku C. Rozbudowane środowisko OntoGen PL zostało przetestowane na zbiorze dokumentów zamieszczonych na stronach intranetowych IŁ. W głównej mierze są to raporty z realizacji prac statutowych obejmujące okres od 2000 do początku 2008 roku. Łącznie było analizowanych 260 opracowań. Biorąc po uwagę stosunkowo długi czas powstawania prac oraz pochodzenie opracowań ze wszystkich komórek badawczych Instytutu, można liczyć na ich reprezentatywność dla rzeczywistych obszarów, w których IŁ jest aktywny. W pierwszym kroku teksty zostały przetworzone, przez konwersję z formatu PDF do postaci czysto tekstowej, redukcję fleksji i inne operacje o charakterze technicznym, związanym z dostosowaniem korpusu do wymagań OntoGen.

Ontologia przedstawiona na rys. 1 była konstruowana metodą grupowania bez nadzoru. W kolejnych krokach procedury aktualna struktura pojęć jest sprowadzana do postaci bardziej szczegółowej, przez podział węzłów odpowiadających pojęciom. Rolą użytkownika, w przypadku takiego podejścia, jest w głównej mierze decyzja co do liczby podwęzłów powstających w wyniku podziału. W trakcie tworzenia struktury dążono do uzyskania jak największej jednorodności zagadnień w poszczególnych węzłach. Najwyższy poziom hierarchii pojęć składa się z trzech podstawowych grup tematycznych, które można hasłowo scharakteryzować jako:

- anteny i sieci radiowe – węzeł 1 (79 dokumentów w tej kategorii),
- urządzenia zasilające sieci radiowe i sieci optyczne – węzeł 2 (96 dokumentów) oraz

- aspekty społeczne i rynkowe telekomunikacji i poczty oraz systemy informacyjne – węzeł 3 (84 skojarzone dokumenty).

Dalszy podział węzła 1 prowadzi do wyodrębnienia zagadnień z dziedziny:

- wzorcowania i pomiarów laboratoryjnych – 1.1,
- badania jakości sieci z wykorzystaniem próbników – 1.2,
- anten i komunikacji radiowej – 1.3.

Węzeł 2 dzieli się na podgrupy tematyczne obejmujące:

- sieci optyczne – 2.1,
- urządzenia zasilające w telekomunikacji – 2.2.

Najbardziej rozbudowaną częścią grafu opisującego strukturę zagadnień badawczych jest podzbiór węzłów zorientowanych wokół problematyki społeczno-ekonomicznej i systemów informacyjnych (główny węzeł 3). Rozbudowa tej części grafu wynika jednak z faktu, że półautomatyczne tworzenie tej ontologii nadzorowała osoba o zainteresowaniach badawczych związanych z tym właśnie obszarem; gdyby to był np. specjalista w zakresie urządzeń zasilających w telekomunikacji, to rozbudowie uległaby niewątpliwie inna część grafu.

Z węzła 3 zostały wyodrębnione grupy zagadnień związanych z:

- rynkiem usług pocztowych – 3.1,
- zastosowaniami teorii gier w telekomunikacji – 3.2,
- problemami regulacji i świadczenia usługi powszechnej – 3.3,
- zarządzaniem wiedzą i społeczeństwem informacyjnym – 3.4,
- rynkiem i ekonomią – 3.5,
- wnioskowaniem opartym o zaawansowaną logikę – 3.6.

Do ciekawych wniosków prowadzi porównanie uzyskanego wyniku ze strukturą organizacyjną IŁ. Stosunkowo łatwo jest skojarzyć część węzłów z jednostkami działającymi w ramach Instytutu, przy czym niektóre z tych skojarzeń wymagają jednak dodatkowej wiedzy na temat profili jednostek (dostępnej np. na stronie internetowej Instytutu <http://www.itl.waw.pl/instytut-struktura-organizacyjna>).

Prototypowe systemy wspomaganie zarządzania wiedzą

Na podstawie wniosków z opisanych działań wstępnych opracowano dwa systemy prototypowe: jeden wspomagający wyszukiwanie informacji w dokumentach elektronicznych opartych na osobistej ontologii użytkownika, drugi wspomagający współdzielenie źródeł wiedzy na podstawie modeli ontologicznych. Funkcje tych systemów połączono następnie w systemie PrOnto.

Wyszukiwanie informacji w dokumentach elektronicznych na podstawie osobistej ontologii użytkownika

Początki prac dotyczących sprawnego wyszukiwania informacji datuje się na lata czterdzieste XX wieku [33]. Wraz z rozwojem internetu nabrały one istotnego praktycznego znaczenia, co potwierdza, że strona wyszukiwarki Google jest najczęściej odwiedzanym miejscem w internecie. W tym miejscu warto zwrócić uwagę na pewną dychotomię związaną z pojęciem wyszukiwania informacji, występuje bowiem:

- wyszukiwanie informacji mającej odpowiedzieć na konkretne pytanie użytkownika, nazwane *szukaniem*,
- wyszukiwanie informacji interesującej użytkownika, ale nie będącej wyrazem jego wprost sprecyzowanych potrzeb, nazwane *przeglądaniem*.

Kierunki te rozwijane są w ramach dziedzin nazwanych w języku angielskim odpowiednio *information retrieval* i *information filtering*. Tradycyjne wyszukiwarki internetowe od samego początku łączyły w pewnym stopniu wsparcie dla obu tych czynności, choć z czasem większą popularność zdobyły rozwiązania lepiej przystosowane do wspomaganie pierwszej z nich, a oparte na wyszukiwaniu dokumentów zawierających podane słowa. Pierwszy nurt podąża w kierunku tzw. *semantycznego internetu* (zob. np. [34]) → formalnego podejścia do opisu informacji opartego na ścisłych definicjach znaczenia gromadzonych danych. Umożliwia to odpowiadanie na pytania przez zastosowanie reguł wnioskowania. Najnowszą praktyczną inkarnacją wyszukiwarek z tego nurtu jest Wolfram Alpha.

Drugi nurt zakłada określenie przez użytkownika jego preferencji, przy czym mogą być one wyrażone wprost lub też mieć postać niejawną, wywnioskowaną z historii jego zachowań. Są to odpowiednio tzw. *content-based information filtering* i *collaborative filtering*.

Jawna postać preferencji może być opisana przez ontologie, przy czym są to najczęściej proste leksykony lub taksonomie. W tym przypadku wyszukiwanie informacji opiera się na analizie treści dokumentów i porównywaniu ich z profilem użytkownika.

Historia zachowań użytkownika, a więc niejawną postać jego preferencji, używana jest do znajdowania podobnie zachowujących się użytkowników i rekomendacji na tej podstawie stosownych produktów (np. filmów, zob. np. [35]). Ten sposób postępowania nabrał w ostatnich latach dużego praktycznego znaczenia, związanego w dużej mierze ze znacznym rozwojem handlu w internecie. Wśród licznych praktycznych implementacji tego pomysłu warto tu wymienić serwis CiteULike, służący wymianie publikacji naukowych.

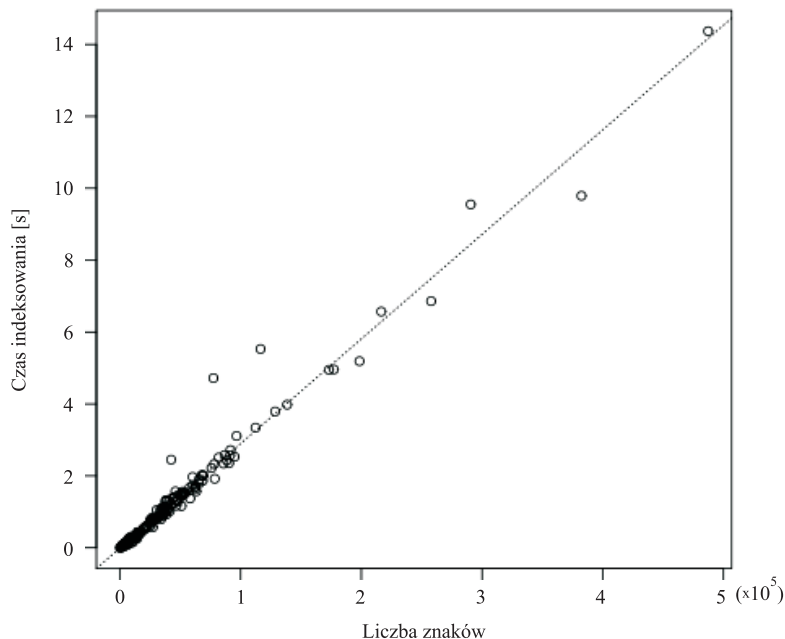
Z ogólnego punktu widzenia, rozróżnienie między *szukaniem* a *przeglądaniem*, przypomina podobny podział obecny w teorii optymalnych decyzji, a mianowicie konieczność wyboru między eksploatacją posiadanej wiedzy a eksploracją w celu pozyskania nowej. Podobnie w ogólnym zadaniu wyszukiwania informacji istotnym problemem jest to, czy sformułowane pytanie jest tym, które użytkownik powinien zadać, aby uzyskać satysfakcjonującą go odpowiedź.

W ostatnim czasie, w wielu wyszukiwarkach pojawiły się zarówno sugestie odnośnie do potencjalnych zapytań, czy ich uszczegółowień, jak i rozszerzanie systemów rekomendacji o informacje strukturalne.

Zdaniem autorów brak jest jednak systemu łączącego w zadowalającym stopniu te zadania, w szczególności przeznaczonego do zastosowań naukowo-badawczych. W pewnym stopniu zadania te są realizowane przez system Ontoshare [36], jednak położono w nim główny nacisk na budowę i wykorzystanie ontologii grupowej, stąd sposób organizacji i wyszukiwania informacji jest w dużym stopniu kwestią indywidualnych preferencji. Takie podejście ogranicza w pewnym stopniu suwerenność użytkownika.

Z tych powodów, w ramach działań grupy tematycznej zajmującej się wykrywaniem wiedzy rozpoczęto prace nad środowiskiem PrOnto, którego głównym celem ma być wspomaganie dzielenia się wiedzą i wyszukiwania potrzebnych informacji w instytucjach naukowo-badawczych. W szczególności wspierany ma być pierwszy krok spirali hermeneutycznej. Reprezentacja preferencji użytkowników została oparta na indywidualnych ontologiach. W tej chwili jest to zbiór pojęć powiązanych ze sobą relacjami. Z każdym pojęciem związany jest zestaw fraz kluczowych z przypisanymi im wagami, określającymi istotność danej frazy kluczowej dla konkretnego pojęcia. Środowisko PrOnto będzie bardziej szczegółowo opisane dalej.

Jednym z problemów szczegółowych, niezbędnych przy rozwoju takich środowisk, jak system PrOnto, jest wybór **algorytmu wyszukiwania fraz kluczowych**. Jednym ze standardowych zadań wyszukiwania wzorców w algorytmice jest problem znany jako *multiple pattern string matching*, a więc problem wyszukiwania słów z zadanego zbioru w łańcuchu znaków. Można łatwo dostosować sformułowanie tego problemu do zadania wyszukiwania fraz kluczowych z zadanego ich zbioru, definiowanego, np. przez struktury ontologiczne użytkowników.



Rys. 2. Zależność czasu indeksowania od rozmiaru dokumentu

Złożoność obliczeniowa trywialnego algorytmu dla tego zadania, używającego operacji wyszukiwania ciągu znaków w tekście dla każdego wzorca jest liniowa względem liczby wzorców, co przy zakładanej ich liczbie (rzędu 10 mln), czyni go nieużywalnym. Dlatego zdecydowano się na implementację algorytmu Aho-Corasick [37], służącego do szybkiego wyszukiwania w tekście łańcuchów znaków pochodzących z zadanego wcześniej zbioru. Istotną cechą tego algorytmu jest jego złożoność obliczeniowa, która nie zależy od liczby wyszukiwanych wzorców. Ze względu na planowany rozmiar systemu, struktury używane do indeksowania przechowywane są w bazie danych. Dzięki temu możliwe jest, np. indeksowanie dokumentów przy użyciu wszystkich tematów z Wikipedii (stąd właśnie oszacowanie rzędu 10 mln fraz kluczowych). Eksperymenty pokazują, że czas indeksowania dokumentu zależy liniowo od jego rozmiaru (rys. 2), a nie od liczby fraz kluczowych, co zgodne jest z oszacowaniem teoretycznym.

Innym ważnym problemem szczegółowym jest **ocena zgodności dokumentu z pojęciem ontologicznym**. W celu dokonania takiej oceny, system został formalnie zdefiniowany jako krotka $(\mathbf{D}, \mathbf{C}, \mathbf{K}, \mathbf{f}, \mathbf{g})$, złożona z:

- zbioru dokumentów $\mathbf{D} = \{d_1, \dots, d_{nd}\}$,
- zbioru pojęć $\mathbf{C} = \{c_1, \dots, c_{nc}\}$,
- zbioru fraz kluczowych $\mathbf{K} = \{k_1, \dots, k_{nk}\}$,

- funkcji określającej wagę danej frazy kluczowej dla danego pojęcia $f: C \times K \rightarrow R$,
- funkcji określającej wyniki indeksowania $g: D \times K \rightarrow R$.

Na tej podstawie można zdefiniować funkcję oceny zgodności dokumentu z pojęciem ontologicznym $h: D \times C \rightarrow R$ w postaci sumy wyników indeksowania ważonej frazami kluczowymi przypisanymi danemu pojęciu:

$$h(d_i, c_j) = \sum_{k_k} f(c_j, k_k) \cdot g(d_i, k_k)$$

Miara ta prezentowana jest w interfejsie użytkownika i używana do porządkowania wyników wyszukiwania.

Współdzielenie źródeł wiedzy na podstawie modeli ontologicznych

Problemy gromadzenia, organizacji i wymiany źródeł wiedzy są już od dłuższego czasu obecne wśród tematów badań w dziedzinie zarządzania informacją i wiedzą. Jednak to ostatnie lata przyniosły rosnące zainteresowanie tymi zagadnieniami, ze względu na zwiększone zapotrzebowanie na metody umożliwiające zapanowanie nad wielością źródeł informacji dostępnych zarówno w sieciach publicznych, jak i wewnątrz organizacji i przedsiębiorstw.

Najbardziej klasycznym sposobem uzyskania dostępu do informacji jest jej *wyszukiwanie według słów kluczowych*. Technika, ulegająca stałemu ulepszaniu i poprawie, zaczyna, wraz ze wzrostem wolumenu informacji i wymagań użytkowników sieci, ocierać się o granicę swoich możliwości. Oczywiście ten sposób poszukiwania interesujących źródeł wiedzy będzie jeszcze przez dłuższy czas dominował, jednak nawet giganci rynku wyszukiwarek są zmuszeni do poszukiwania nowych rozwiązań usprawniających procesy poszukiwania i porządkowania.

Wyszukiwarka Wolfram Alpha interpretuje słowa kluczowe wchodzące w skład zapytania użytkownika i prezentuje wyniki w uporządkowanej formie, Google Squared zaś organizuje rezultaty wyszukiwania w formie tabelarycznej, przy czym kolejne wiersze odpowiadają znalezionym obiektom, kolumny zaś to ich atrybuty. Należy zwrócić uwagę na aktualność tych rozwiązań, jako że zarówno produkt Google, jak i wyszukiwarka Stephena Wolframa^① zostały udostępnione w 2009 roku.

Innym sposobem na poprawę wyników wyszukiwania źródeł informacji i wiedzy jest personalizacja, polegająca na umożliwieniu użytkownikom modyfikacji uzyskanych w sposób automatyczny rezultatów i przypisywaniu im komentarzy. Przykładem tego typu rozwiązania jest usługa Google SearchWiki.

Portale społecznościowe, których rozkwit nastąpił w ostatnich latach, gromadzą informację o swoich użytkownikach, ich zainteresowaniach i aktywności. Pojawiają się pomysły wykorzystania struktury takich portali i powiązań między ich użytkownikami, jako dodatkowego wymiaru organizującego wiedzę i pozwalającego uzyskiwać lepsze wyniki wyszukiwania [38].

Szczególnym zagadnieniem jest organizacja procesów gromadzenia i zarządzania zasobami wiedzy w organizacjach badawczych, bądź w tzw. *Communities of Practice* (CoP), w których szczególny nacisk kładzie się na wymianę informacji niezbędnej do rozwiązywania konkretnych problemów, tworzenie nowej wiedzy i wsparcie kreatywności [5], [8]. Znane są też systemy zorganizowane w postaci *sieci społecznych*, wspomagające wymianę informacji naukowej i ułatwiające kontakt osób zaangażowanych w działalność badawczą, w szczególności można przywołać IndexCopernicus, BioMedExperts, czy BioCrowd.

^① Stephen Wolfram jest również znany jako autor programu Mathematica.

W większości *sieci społecznych* informacja jest organizowana za pomocą tzw. tagów, czyli słów kluczowych, które są przypisywane przez użytkowników zasobom, co lokuje ten sposób indeksowania na przeciwnym biegunie w stosunku do sposobu indeksowania stosowanego w wyszukiwarkach internetowych, które wiążą słowa kluczowe z dokumentami na podstawie zawartości tych dokumentów^①.

W ramach projektu On-To-Knowledge powstał system OntoShare [36], którego zadaniem jest umożliwienie wymiany źródeł wiedzy w ramach CoP. Podstawowym założeniem systemu jest wykorzystanie wspólnej dla CoP ontologii. Użytkownik wyraża chęć udostępnienia dokumentu w OntoShare. Dokument jest analizowany przez system i opisywany zestawem słów kluczowych. Każdy z użytkowników systemu jest charakteryzowany przez profil, złożony ze zbioru właściwych jego sposobowi percepcji dziedziny pojęć ontologicznych, a każde pojęcie jest z kolei powiązane ze zbiorem słów kluczowych. OntoShare bada podobieństwo pojęć tworzących profile użytkowników do opisu dokumentu i, w przypadku stwierdzenia znacznej zbieżności, proponuje zaetykietowanie dokumentu sygnaturą jednego lub wielu pojęć. Użytkownicy mogą również dodawać nowe pojęcia do swoich profili. Dzięki takiemu modelowi opisu wiedzy, można lokalizować we wspólnej przestrzeni najbardziej interesujące z punktu widzenia użytkownika dokumenty, ale także i osoby deklarujące znajomość określonej tematyki.

Nieco innym podejściem do problemu wymiany wiedzy opierającego się na opisie ontologicznym jest projekt SWAP (*Semantic Web and Peer-to-Peer*, [39], [40]). Projekt powstał w wyniku obserwacji gwałtownego rozwoju sieci bezpośredniej wymiany^② (P2P) i miał na celu stworzenie sieci P2P umożliwiającej wymianę wiedzy między niezależnymi agentami, posługującymi się wspólnymi, lub prywatnymi ontologiami^③. Zakłada się, że wiedza jest rozproszona między węzłami sieci P2P i podstawowym zadaniem systemu jest dopasowywanie opisujących ją ontologii poszczególnych węzłów, w celu lokalizacji źródeł odpowiadających w największym stopniu zapytaniom użytkowników. Szczególnie istotne są w tym kontekście zagadnienia związane z dopasowywaniem ontologii (*Ontology Matching*, zob. [41]), a metody dopasowania stanowią znaczną część prac badawczych w ramach projektu SWAP. Jednym z produktów projektu jest Bibster [42], system do wymiany informacji bibliograficznej w środowisku rozproszonym.

Spośród różnych innych prób zmierzania się z zagadnieniem zarządzania wiedzą na podstawie modeli ontologicznych, warto także wspomnieć o projekcie SEKT (*Semantically – Enabled Knowledge Technologies*), którego tematyka obejmuje szereg zagadnień związanych z tworzeniem i dopasowywaniem ontologii oraz wymianą wiedzy.

Co ciekawe, wyniki prac nad porządkowaniem wiedzy na podstawie modeli ontologicznych, są dziś trudno dostępne w domenie publicznej^④. Można podejrzewać, że rezultaty okazały się niewystarczająco zadowalające i nie przetrwały próby czasu. Z drugiej jednak strony, echa tych projektów można znaleźć na stronach internetowych komercyjnych partnerów tworzących projektowe konsorcja. Informacja na tych stronach ma charakter przeważnie marketingowy, jednak można na jej podstawie przypuszczać, że firmy wykorzystały przynajmniej wnioski z prac badawczych przy tworzeniu portfolio swoich produktów.

^① Skoncentrowano się tu na wyszukiwarkach operujących na dokumentach tekstowych.

^② Sieci P2P (*Peer-to-Peer Networks*), takie jak BitTorrent, Gnutella, Napster, eDonkey, Kad, służą dziś przede wszystkim do bezpośredniej wymiany plików. Istnieją także inne zastosowania sieci P2P, w tym również wymiana metadanych tworzących sieci semantyczne, np. sieć EduTella.

^③ Indywidualne ontologie nazwane zostały kontekstami (*contexts*), nazwa ontologia zaś jest wtedy zarezerwowana dla współdzielonych struktur semantycznych.

^④ Łatwo dostępny jest tylko Bibster (<http://bibster.semanticweb.org>).

System PrOnto

W przekonaniu autorów niniejszego opracowania, organizacja wiedzy w strukturze ontologicznej sprawdza się znacznie lepiej jako podstawa reprezentacji wiedzy w systemie jej współdzielenia niż w przypadku oparcia się jedynie na słowach kluczowych. Pojęcia ontologiczne i relacje między nimi pozwalają systematyzować wiedzę w intuicyjny sposób, odpowiadający spojrzeniu na zagadnienia mieszczące się w obszarze zainteresowań osoby lub grupy osób. Uzyskane w ten sposób indywidualne profile mogą być w przyszłości wykorzystane przy tworzeniu mechanizmów współdzielenia wiedzy.

Modele ontologiczne są również istotne w kontekście poznawczym, związanym z poszerzaniem osobistego horyzontu hermeneutycznego, ułatwiając lokalizację pojęć leżących na granicy horyzontów indywidualnych i nawigację w obrębie sąsiednich, „przylegających” horyzontów. Prosty przykład użyteczności takiego rozwiązania to przyjęcie punktu widzenia osoby, która wykazuje się znaczną wiedzą w określonej tematyce i wykorzystanie jej struktury ontologicznej jako mapy prowadzącej przez dany obszar zagadnień, ułatwiającej przyswajanie nowej wiedzy w sposób uporządkowany.

Frazy kluczowe są w takich zastosowaniach bardziej problematyczne. Wyszukiwanie informacji na podstawie fraz kluczowych wymaga dokładnej znajomości tych fraz. Ontologie stanowią natomiast wyższy poziom abstrakcji, bliższy sposobowi opisu dziedziny właściwemu ludziom. Dostęp do wielu profili użytkowników, na poziomie opisu ontologicznego, daje możliwość odnalezienia właściwych fraz kluczowych, powiązanych z pojęciami tworzącymi te profile, co z kolei zwiększa znacznie prawdopodobieństwo dotarcia do istotnej informacji, nawet w przypadku nikłej znajomości dziedziny, w tym fraz kluczowych charakteryzujących opisujące ją dokumenty, co jest przypadkiem nagminnym w procesie przyswajania nowej wiedzy.

Wspomniany system PrOnto jest środowiskiem do organizacji i współdzielenia źródeł wiedzy, zorientowanym na wspieranie procesów poznawczych i kreatywności w środowisku badawczym. Został stworzony w Instytucie Łączności i istnieje obecnie w fazie prototypu. Dostęp do systemu odbywa się za pośrednictwem przeglądarki www i jest chroniony hasłem. Użytkownik dostarcza do wspólnej bazy dokumenty, które uważa za istotne dla prowadzonych przez siebie badań. Opisuje jednocześnie swój obszar zainteresowań za pomocą struktury ontologicznej, bliskiej mapie pojęć, a także proponuje słowa kluczowe charakteryzujące dostarczone przez siebie dokumenty i wiąże je z elementami swojej ontologii. Ontologia jest następnie używana jako perspektywa, służąca do porządkowania dokumentów względem rankingu słów kluczowych powiązanych z pojęciami, a także do poszukiwania profili ontologicznych posługujących się podobnymi znaczeniowo pojęciami. Zaimplementowano szereg narzędzi wspierających zarówno refleksję hermeneutyczną nad indywidualną strukturą pojęć, jak i współdzielenie źródeł wiedzy.

Model ontologiczny

Reprezentacja źródeł wiedzy w środowisku PrOnto (rys. 3) jest tworzona na podstawie schematu, który można w sposób symboliczny przedstawić jako strukturę:

$$KR := (H, C, R, K, D, \alpha_C, \alpha_R, \delta, \gamma_C, \gamma_D)$$

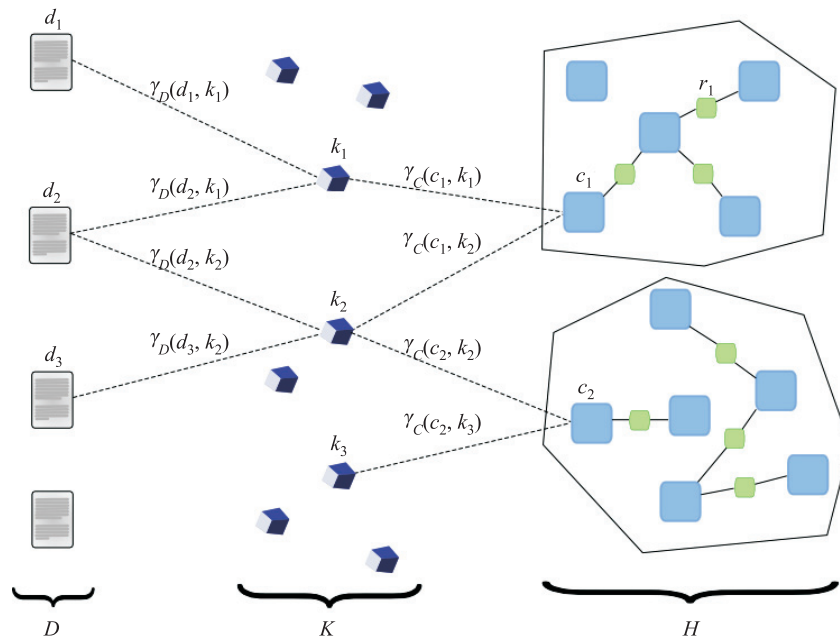
gdzie:

- H jest zbiorem horyzontów albo perspektyw odpowiadających indywidualnym lub grupowym punktom spojrzenia na zasoby wiedzy przechowywane w systemie; zdefiniowany w ten sposób horyzont stanowi uogólnienie użytkownika systemu,

- C jest zbiorem pojęć tworzących horyzont H ,
- R jest zbiorem relacji między pojęciami z C ,
- $\alpha_C : C \rightarrow H$ i $\alpha_R : R \rightarrow H$ stanowią, odpowiednio, przypisanie pojęć z C i relacji z R do horyzontów z H ,
- $\delta : R \rightarrow C \times C$ przypisuje relacjom z R pojęcia będące jej argumentami,
- K jest zbiorem fraz lub słów kluczowych,
- D jest zbiorem bądź bazą dokumentów,
- $\gamma_D : D \times K \rightarrow \mathfrak{R}$ jest miarą reprezentatywności fraz z K dla dokumentów z D ; została zaimplementowana klasyczna miara TF-IDF (*Term Frequency – Inverse Document Frequency*), wartość $\text{TF-IDF}(k \in K, d \in D)$ rośnie wraz z częstością występowania frazy k w dokumencie d i maleje wraz z całkowitą liczbą dokumentów zawierających k ,
- $\gamma_C : C \times K \rightarrow \mathfrak{R}$ jest miarą będącą subiektywną oceną rozkładu prawdopodobieństwa $P(c | k), \forall c \in C \forall k \in K$; jest to miara definiowana dla każdego pojęcia $\forall c \in C$ przez właściciela horyzontu H , do którego c należy; każdemu pojęciu z C może zostać przypisana dowolna liczba fraz kluczowych z K .

Na podstawie przedstawionego wyżej modelu została stworzona podstawowa dla systemu funkcja rankingowa (analogiczna do stosowanej funkcji oceny zgodności dokumentu z pojęciem ontologicznym $\mathbf{h: D \times C \rightarrow R}$), która umożliwi porządkowanie dokumentów względem struktury pojęć:

$$\phi(c, d) = \sum_{k \in K} \gamma_C(c, k) \cdot \gamma_D(d, k) \quad \forall d \in D, \forall c \in C$$

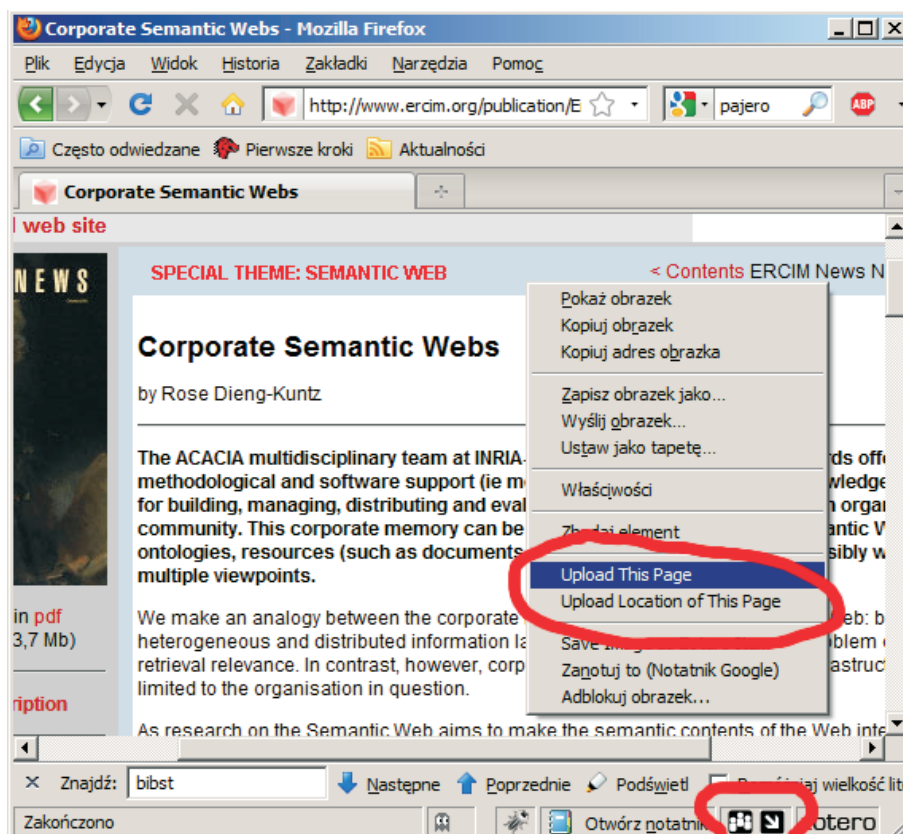


Rys. 3. Reprezentacja źródeł wiedzy w PrOnto

Współdzielenie wiedzy

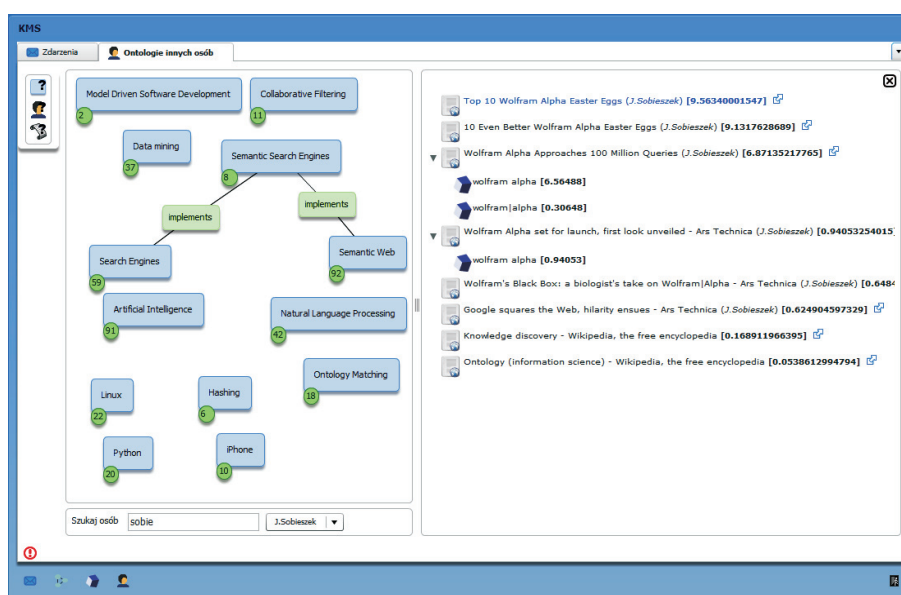
Dalej zostaną pokrótce przedstawione mechanizmy dzielenia wiedzy w systemie PrOnto. Stanowią one, obok indywidualnego porządkowania źródeł wiedzy, jedną z dwóch zasadniczych funkcjonalności prototypowej wersji systemu. Perspektywa indywidualna nie będzie tutaj przedstawiona.

Dostarczanie źródeł wiedzy. Jak to już sygnalizowano wcześniej, poszukiwanie źródeł wiedzy związanych z przedmiotem badań jest istotnym elementem procesu kreowania wiedzy i znalazło odzwierciedlenie w hermeneutycznej spirali EAIR (*Enlightenment-Analysis-Hermeneutic Immersion-Reflection*), jako krok analizy (Analysis), [5]. Podczas procesu poszukiwań źródeł, część z nich zostaje odrzucona, część zaś zaklasyfikowana jako szczególnie istotne w kontekście badanego zagadnienia. PrOnto umożliwia umieszczenie dokumentów w dzielonej między swoich użytkowników wspólnej przestrzeni źródeł wiedzy (w przedstawionym modelu odpowiada on bazie dokumentów D). Aby dodatkowy nakład pracy związany z rejestrowaniem dokumentu w systemie był jak najmniejszy, została stworzona dedykowana „wtyczka” do przeglądarki Firefox, która po zainstalowaniu sprowadza całą czynność do prostych operacji wykonywanych za pomocą myszy (rys. 4).



Rys. 4. Udostępnianie strony WWW. Zaznaczone zostały elementy powiązane z PrOnto Firefox Extension

Perspektywy innych użytkowników systemu. System PrOnto umożliwia spojrzenie na zgromadzone dokumenty z perspektywy innych użytkowników, a ściślej mówiąc, przez pryzmat ich struktury ontologicznej. Funkcjonalność taka daje możliwość poszerzenia własnego horyzontu pojęciowego w sposób uporządkowany, a szczególnie przydatna wydaje się na etapie poznawania dziedziny, w której właściciel horyzontu jest biegły. Ekran aplikacji został przedstawiony na rys. 5. Po lewej jego stronie zobrazowana została mapa pojęć, po prawej lista dokumentów odpowiadających pojęciu Semantic Search Engines, uporządkowana według wartości funkcji rankingowej ϕ . Pod ikoną dokumentu wyświetlone zostały frazy kluczowe, które są wspólne dla pojęcia i dokumentu, wraz z przypisanymi im wartościami $\gamma_c(c, k) \cdot \gamma_D(d, k)$ ^①.



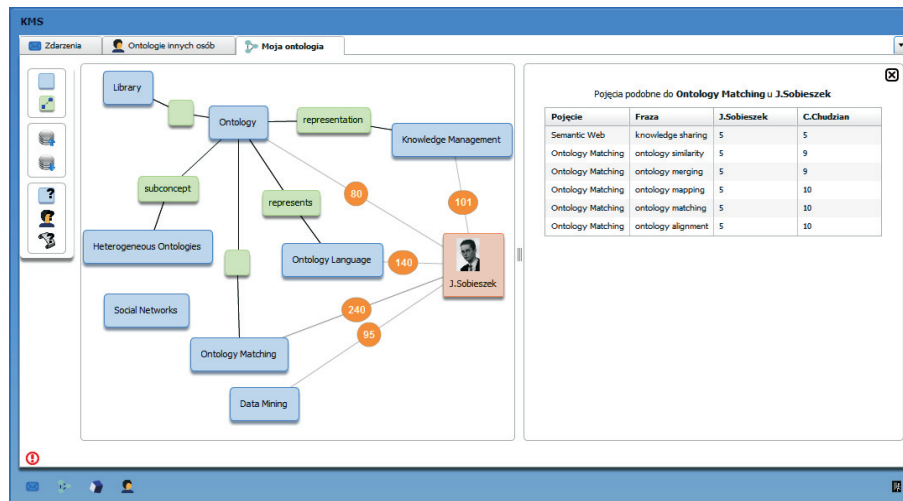
Rys. 5. Dostęp do dokumentów z perspektywy profili innych użytkowników

Dopasowanie ontologii. Kolejna klasa analiz, oferowana przez PrOnto, to dopasowywanie ontologii. Użytkownik uzyskuje dzięki temu informację o pojęciach występujących w innych profilach, które są podobne do pojęć, którymi on sam się posługuje. Obecna wersja PrOnto dopasowuje pojęcia według równości przypisanych im etykiet, oraz przez porównanie zbiorów słów kluczowych przypisanych pojęciom przez użytkowników (rys. 6). Dla pojęć, które mają swoje odpowiedniki w profilach innych użytkowników istnieje możliwość śledzenia relacji (rys. 7). Przedstawiane są, w formie listy, pojęcia sąsiednie do dopasowanych pojęć z innych ontologii, przy czym sąsiedniość oznacza tu bycie w relacji z pojęciem.

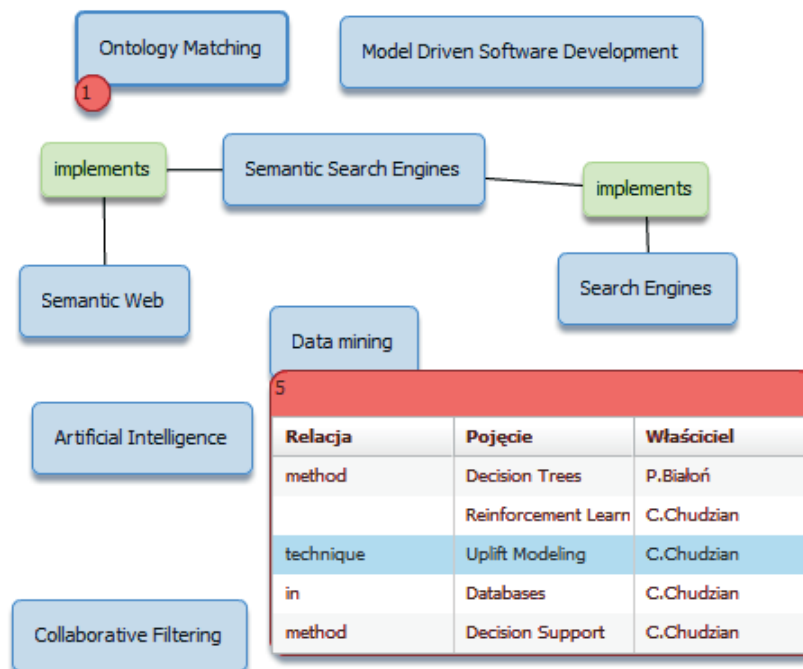
Wielowymiarowe wyszukiwanie dokumentów. PrOnto jest wyposażone w zaawansowaną wyszukiwarkę osób, pojęć, dokumentów i fraz kluczowych (rys. 8), która prezentuje wyniki poszukiwań w wielowymiarowej strukturze. Na rys. 8, w prawej jego części, widać wynik wyszukiwania dokumentów, którym przypisano najbardziej dopasowane do nich pojęcia, przy czym miara dopasowania

^① d – ustalony dokument, c – pojęcie, k – ustalona fraza kluczowa.

to znana z wcześniejszych rozważań funkcja rankingowa ϕ . Pojęcia pochodzą z profili wszystkich użytkowników. Po rozwinięciu ikony pojęcia są wyświetlane frazy kluczowe, które charakteryzują jednocześnie dokumenty i pojęcia, determinując ich pozycję w rankingu. Lewa część okna z rys. 8 stanowi interfejs służący do definiowania powiązań pojęć ze słowami kluczowymi, w tym wartości miary γ_C .



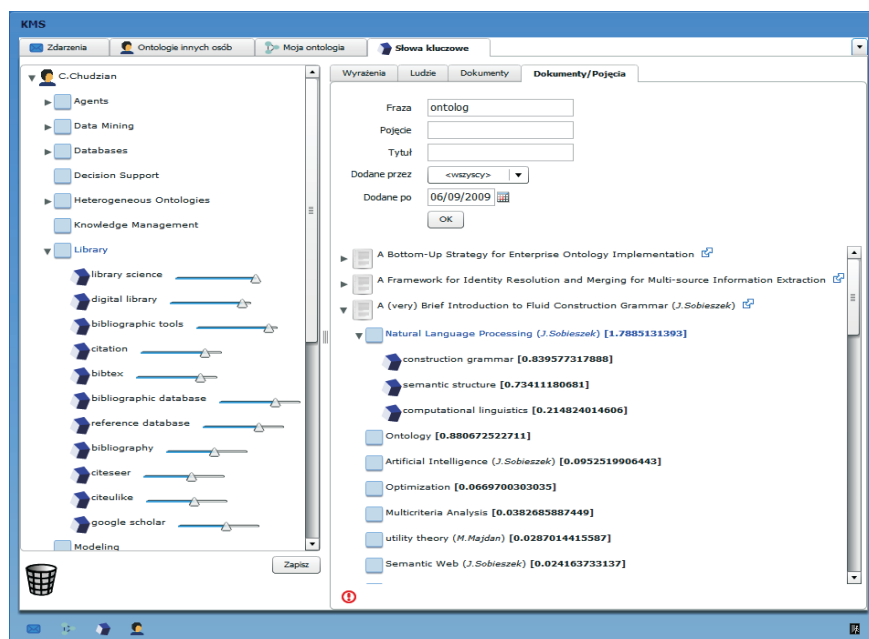
Rys. 6. Podobieństwo profili użytkowników



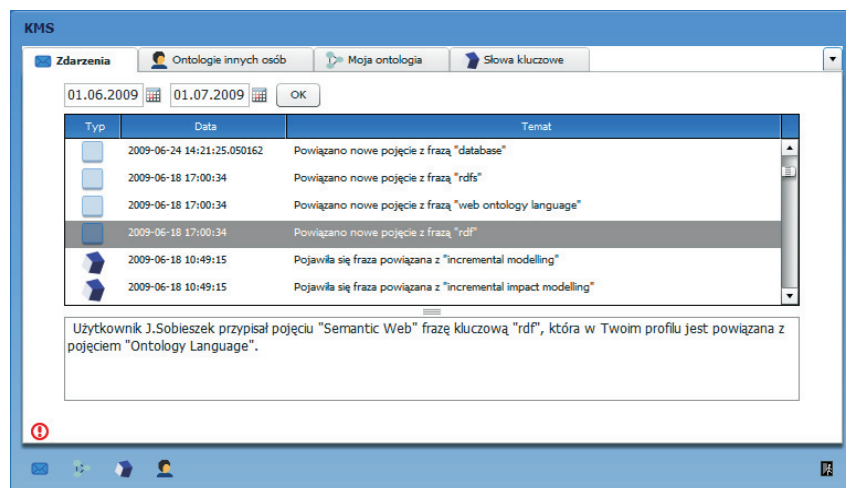
Rys. 7. Śledzenie relacji

Wyniki analiz, w trybie off-line

Oprócz interaktywnego trybu pracy, PrOnto wykonuje analizy zgromadzonej wiedzy bez udziału użytkownika. Wyniki tych analiz, prezentowane w formie listy wiadomości, mają wspomagać użytkownika w konstruowaniu własnego profilu i informować o istotnych zdarzeniach w systemie (rys. 9). Z punktu widzenia współdzielenia wiedzy ważne jest szczególnie dostarczanie informacji o nowych pojęciach, pokrewnych własnym, które pojawiają się w profilach innych użytkowników i informacja o dostarczeniu do systemu źródeł wiedzy leżących w obszarze zainteresowania użytkownika.



Rys. 8. Dokumenty, pojęcia i frazy kluczowe



Rys. 9. Informacja o zdarzeniach

Zagadnienia implementacyjne

Środowisko PrOnto zostało w całości oparte na oprogramowaniu typu *open source*. Należy jednak nadmienić, że wykorzystane technologie są już obecnie szeroko stosowane w aplikacjach wymagających zapewnienia niezawodności i wysokiej jakości rozwiązania, w tym również w oprogramowaniu komercyjnym. Podstawowe rozwiązania technologiczne użyte przy implementacji PrOnto, to relacyjna baza danych PostgreSQL, środowisko do tworzenia aplikacji www (tzw. *Web Application Framework*) Django, oparte na języku skryptowym Python oraz środowisko Adobe Flex do tworzenia aplikacji Flash. Ponadto w warstwie oprogramowania realizującego indeksowanie dokumentów frazami kluczowymi, PrOnto opiera się na autorskich kodach napisanych w języku C.

Podsumowanie

W ramach projektu powstał prototyp systemu do organizacji i współdzielenia źródeł wiedzy – PrOnto. W przypadku tego typu oprogramowania, działającego na podstawie idei sieci społecznych, ocena rezultatów jest możliwa dopiero po osiągnięciu przez system masy krytycznej, mierzonej liczbą użytkowników. Zostały też przygotowane zintegrowane mechanizmy ankietowania użytkowników, dostępne z poziomu aplikacji PrOnto.

Planowany jest rozwój PrOnto w kilku kierunkach, z których najważniejsze to:

- Zwiększenie ergonomii interfejsu użytkownika. Etap ten będzie następstwem analizy uwag zgłaszanych przez użytkowników w trakcie prowadzenia końcowych testów.
- Rozszerzenie zakresu automatycznej analizy źródeł wiedzy i zestawu narzędzi wspomagających tworzenie indywidualnej ontologii.
- Rozbudowa metod dopasowywania ontologii.
- Wzbogacenie profili ontologicznych, przez wzmocnienie ich semantyki. Konieczne jest poprawienie siły wyrazu struktur pojęciowych, przez zwiększenie formalizmu ich opisu.
- Większa skalowalność rozwiązania. Przy rosnącej liczbie użytkowników niezbędne będzie dostosowanie znacznej części mechanizmów funkcjonujących obecnie na zasadzie prototypów.
- Opracowanie nowych procedur i funkcji rankingowych dla porządkowania i współdzielenia źródeł wiedzy.
- Tworzenie ścieżek poznawczych (*learning paths*). Ścieżka taka, w przypadku PrOnto, przechwuje ślad procesu poznawczego użytkownika, w postaci kolejno dostarczanych dokumentów, tworzonych pojęć i przypisywanych im słów kluczowych. Oczekuje się, że odtworzenie takiej ścieżki w przyszłości, przez innych użytkowników, może w znacznym stopniu usprawnić zdobywanie przez nich wiedzy.

Wnioski z dotychczasowych prac

Przedstawione w tym artykule rezultaty prac grupy tematycznej zajmującej się zagadnieniem wykrywania wiedzy (dokładniej, grupy i: *Systemy wspomagania decyzji regulacyjnych: Wykrywanie wiedzy w dużych zbiorach danych telekomunikacyjnych. PBZ Usługi i sieci teleinformatyczne następnej generacji – aspekty techniczne, aplikacyjne i rynkowe*) wskazują na duże możliwości przyszłych zastosowań inżynierii ontologicznej we wspomaganiu prac badawczych, zwłaszcza w telekomunikacji, a także prac regulacyjnych.

Wspomaganie takie nie może być jednak nastawione na pełną automatyzację, musi opierać się na dobrym zrozumieniu rozróżnienia między wiedzą jawną a ukrytą, doceniać rolę użytkownika-

eksperta w procesach kreowania wiedzy, doceniać też różnice między akademickimi procesami kreowania wiedzy a organizacyjnymi procesami tego typu.

W pracach grupy tematycznej wykonano dwa prototypowe systemy wspomaganie prac badawczych, a także uzyskano wiele rezultatów i przemyśleń metodologicznych i teoretycznych, wykonano też wstępne testowanie systemów pilotowych. Jednakże ograniczony czas prac projektu ograniczył też zakres testowania, które powinno w przyszłości obejmować zarówno wybrane zespoły badawcze i jednostki funkcjonalne IŁ lub współpracujących uczelni, jak i zarządy – dyrekcje lub rektoraty tych instytucji.

Jest uzasadnione oczekiwanie, że uzyskane rezultaty mogą mieć duże znaczenie także w innych zastosowaniach np. w telekomunikacji – w pracach urzędów regulacyjnych czy też w rozwoju Internetu Przyszłości w związku z trendem do uwzględnienia świadomości zawartości przekazu w zarządzaniu siecią.

Bibliografia

- [1] Chudzian C. i inni: *Wykrywanie wiedzy w dużych zbiorach danych: Analiza tekstu i inżynieria ontologiczna*. Sprawozdanie PBZ “Usługi i sieci teleinformatyczne następnej generacji – aspekty techniczne, aplikacyjne i rynkowe”. Grupa tematyczna i: *Systemy wspomaganie decyzji regulacyjnych: Wykrywanie wiedzy w dużych zbiorach danych telekomunikacyjnych*. Warszawa, Instytut Łączności, 2009
- [2] Kant I.: *Kritik der Reinen Vernunft*. 1781. Polskie tłumaczenie: *Krytyka czystego rozumu*. Warszawa, PWN, 1957
- [3] Król Z.: The emergence of new concepts in science. In: Wierzbicki A.P., Nakamori Y., (eds): *Creative Environments*, op. cit., 2007
- [4] Jung C.G.: *Typy psychologiczne*. Warszawa, Wydawnictwo KR, 2009
- [5] Wierzbicki A.P., Nakamori Y.: *Creative Space: Models of Creative Processes for the Knowledge Civilization Age*. Vol. 10 of *Studies in Computational Intelligence*. Berlin, Springer, 2006
- [6] Polanyi M.: *The Tacit Dimension*. London, Routledge and Kegan, 1966
- [7] Nonaka I., Takeuchi H.: *The Knowledge-Creating Company. How Japanese Companies Create the Dynamics of Innovation*. New York, Oxford University Press, 1995. Polskie tłumaczenie: *Kreowanie wiedzy w organizacji*. Poltext, 2000
- [8] Wierzbicki A.P., Nakamori Y. (eds): *Creative Environments: Issues of Creativity Support for the Knowledge Civilization Age*. Vol. 59 of *Studies in Computational Intelligence*. Berlin, Springer, 2007
- [9] Wierzbicki A.P., Nakamori Y.: *Knowledge sciences – some new developments*. *Zeitschrift für Betriebswirtschaft*, vol. 77, no 3, 2007, pp. 271-296
- [10] Bishop C.M.: *Pattern Recognition and Machine Learning*. Signapore, Springer, 2006
- [11] Gadamer H-G.: *Warheit und Methode. Grundzüge einer Philosophischen Hermeneutik*. Tübingen, J.B.C. Mohr (Siebeck), 1960
- [12] Ren H., at al.: *Electronic support for knowledge creation in a research institute*, *Journal of Systems Science and Systems Engineering*, vol. 16, no 2, 2007

- [13] Davenport T., Prusak L.: *Working Knowledge: How Organizations Manage what they Know*. Boston Ma., Harvard Business School Press, 1998
- [14] Dieng R., Corby O.: *Knowledge Engineering and Knowledge Management: Methods, Models and Tools*. Berlin-Heidelberg, Springer Verlag, 2000
- [15] Pfeffer J., Sutton R.I.: *The Knowing – Doing Gap: How Smart Companies Turn Knowledge into Action*. Boston, MA., Harvard Business School Press, 2000
- [16] Ren H. at al.: Ontology construction and its applications in local research communities. In: Granat J., Dolk D. (eds) *Modelling and Decision Support for Network-based Services*, (in print)
- [17] Nonaka I.: *The knowledge creating company*. Harvard Business Review, no. 69, 1991, pp. 96–104
- [18] Toffler A.: *Future Shock*. New York, Random House, 1970
- [19] Akkermans H., Gordijn J.: Ontology engineering, scientific method and the research agenda. In: Motta E., at al. (eds): *Managing Knowledge in a World of Networks*. Berlin-Heidelberg, Springer, 2006, pp. 112-125
- [20] Heidegger M.: *Sein und Zeit*. Halle, Niemayer, 1927
- [21] Mizoguchi R., at al.: *Construction and deployment of a plant ontology*. In: Dieng R., Corby O. (eds), *Knowledge Engineering and Knowledge Management*, 12th International Conference, EKAW 2000, Juan-les-Pin, 2000, pp.113-128
- [22] Corcho O., Fernández-López M., Gómez-Pérez A.: *Methodologies, tools and languages for building ontologies: Where is their meeting Point?*, Data & Knowledge Engineering, no. 46, 2003, pp. 41-64
- [23] Pinto H.S., Martins J.P.: *Ontologies: How can they be built?* Knowledge and Information Systems, no. 6, 2004, pp. 441-464
- [24] Bontas E.P., Tempich C.: Ontology engineering: A reality check. In: Meersman R. at al. (eds), *The 5th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE2006)*, vol. 4275 of Lecture Notes in Computer Science, Montpellier, France, Springer, 2006, pp. 836-854
- [25] Gómez-Pérez A., Fernández-López M., Corcho O.: *Ontological Engineering*. Springer, 2003
- [26] Curtis J., Baxter D., Cabral J.: *On the application of the cyc ontology to sense disambiguation*. In: Proceedings of the Nineteenth International FLAIRS Conference, Melbourne Beach, FL, 2006, pp. 652-657
- [27] Wierzbicki A.P.: *On the role of intuition in decision making and some ways of multicriteria aid of intuition*. Multiple Criteria Decision Making, no. 6, 1997, pp.65-78
- [28] Wierzbicki A.P.: Intuicja z perspektywy technicznej: znaczenie zasady multimedialnej i zasady emergencji. W: Motycka A. (red): *Wiedza a intuicja*. Warszawa, Wydawnictwo IFiS PAN, 2008, str. 231-264
- [29] Schmidt A., Braun S.: *Context-aware workplace learning support: Concept, experiences and remaining challenges*. In: Proceedings of the First European Conference on Technology-Enhanced Learning (ECTEL 06), Springer, 2006
- [30] Chudzian C.: *Ontology creation process in knowledge management support system for a research institute*. Journal of Telecommunications and Information Technology, no. 4, 2008, pp. 47-53

- [31] Sobieszek J.: *Towards a unified architecture of knowledge management system for a research institute*. Journal of Telecommunications and Information Technology, no. 4, 2008, pp. 54-59
- [32] H. Ren, at al.: *Ontology construction and its applications to a research program*. To appear in Journal of Integrated Computer Aided Engineering
- [33] Bush V.: *As we may think*. The Atlantic Monthly, vol. 176, 1945, pp. 101-108
- [34] Antoniou G., van Harmelen F.: *A Semantic Web Primer*. Cambridge MA., The MIT Press, 2004
- [35] Goldberg D., at al.: *Using collaborative filtering to weave an information tapestry*. Communications of the ACM, no.35, 1992, pp. 61-70
- [36] Davies J., Duke A., Sure Y.: *Ontoshare - an ontology-based knowledge sharing system for virtual communities of practice*. Journal of Universal Computer Science, no. 10(3), 2004, pp.262–283
- [37] Aho A.V., Corasick M.J.: *Efficient string matching: An aid to bibliographic search*. Communications of the ACM, no. 18, 1975, pp. 333–340
- [38] Vogelstein F.: *Great Wall of Facebook: The Social Network's Plan to Dominate the Internet – and Keep Google Out*. Wired Magazine. Issue 17.07, 2009. [Online]. Available <http://www.wired.com>
- [39] Ehrig M., at al.: *SWAP: Semantic Web and Peer-to-Peer Project Deliverable 3.6 Refined Methods*. 2003
- [40] Ehrig M., Tempich C., Aleksovski Z.: *SWAP: Semantic Web and Peer-to-Peer Project Deliverable 4.7 Final Tools*. 2004
- [41] Ehrig M.: *Ontology Alignment: Bridging the Semantic Gap (Semantic Web and Beyond)*. New York, Springer, 2006
- [42] Broekstra J., at al.:(2004) *Bibster - A semantics-based bibliographic peer-to-peer system*. In: Proceedings of the Third International Semantic Web Conference, Hiroshima, 2004, pp. 122-136

Cezary Chudzian



Mgr inż. Cezary Chudzian – absolwent Wydziału Elektroniki i Technik Informatycznych Politechniki Warszawskiej; pracownik Instytutu Łączności (2002-2010); zainteresowania badawcze: zarządzanie informacją i wiedzą z zastosowaniem inżynierii ontologicznej, odkrywanie wiedzy w dużych zbiorach danych, uczenie się maszyn, zaawansowane zagadnienia inżynierii oprogramowania.

e-mail: cchudzia@orange.pl

Janusz Granat



Dr inż. Janusz Granat (1961) – absolwent Wydziału Elektroniki Politechniki Warszawskiej (1987); nauczyciel akademicki i pracownik naukowy Politechniki Warszawskiej (od 1988); pracownik naukowy Instytutu Łączności (od 1997), kierownik Zakładu Zaawansowanych Technik Informatycznych IŁ; przewodniczący grupy badawczej IFIP WG7.6, współpracownik IIASA w Austrii oraz JAIST w Japonii; współorganizator międzynarodowej konferencji *Decision Support for Telecommunications and Information Society (DSTIS)*; kierownik wielu projektów badawczych; autor i współautor wielu publikacji naukowych; zainteresowania naukowe: teoria i metodologia optymalizacji wektorowej, hurtownie danych, analiza danych, projektowanie komputerowych systemów wspomagania decyzji.

e-mail: J.Granat@itl.waw.pl

Edward Klimasara



Mgr Edward Klimasara – absolwent Wydziału Matematyki i Mechaniki Uniwersytetu Warszawskiego (1977); pracownik Instytutu Łączności w Warszawie (od 1984); starszy specjalista w Zakładzie Zaawansowanych Technik Informatycznych; autor i współautor prac z obszaru informatyki i telekomunikacji; zainteresowania zawodowe: zarządzanie wiedzą, zastosowanie technik informatycznych w telekomunikacji, medycynie, transporcie, administracji i edukacji.

e-mail: E.Klimasara@itl.waw.pl

Andrzej P. Wierzbicki



Prof. zw. dr hab. inż. Andrzej P. Wierzbicki (1937) – absolwent Wydziału Łączności Politechniki Warszawskiej (1960); nauczyciel akademicki i pracownik naukowy wielu uczelni (Politechnika Warszawska, Uniwersytet Minnesota, Uniwersytet Browna, Uniwersytet Kioto, IIASA, JAIST), pracownik naukowy Instytutu Łączności w Warszawie (od 1996), dyrektor naczelny IŁ (1996–2004); organizator działalności badawczej i naukowej; członek Information Society Advisory Group (ISTAG) powołany przez Komisję Europejską, przewodniczący Zespołu Doradców KBN ds. Naukowej Współpracy Międzynarodowej, wiceprzewodniczący Komitetu Prognoz „Polska 2000 Plus” przy Prezydium PAN; autor licznych publikacji; zainteresowania naukowe: teoria i metodologia optymalizacji wektorowej, wspomaganie decyzji i projektowania, teoria oraz metody obliczeniowe optymalizacji, techniki i sztuka negocjacji, zjawiska cywilizacyjne, rynkowe oraz techniczne związane z pojęciami społeczeństwa informacyjnego i cywilizacji informacyjnej.

e-mail: A.Wierzbicki@itl.waw.pl

Jarosław Sobieszek



Mgr inż. Jarosław Sobieszek – absolwent wydziału Elektroniki i Technik Informatycznych Politechniki Warszawskiej; pracownik Instytutu Łączności (2003-2010); zainteresowania badawcze: zarządzanie wiedzą, uczenie się maszyn, sztuczna inteligencja, oparte na modelach podejście do rozwoju oprogramowania.

e-mail: j.sobieszek@neo.pl