

*WOJCIECH FOLTA\**

*ANNA STOLIŃSKA\**

Uniwersytet Pedagogiczny w Krakowie

## **BUDOWA SIECI SEMANTYCZNEJ – NOWA PERSPEKTYWA**

### **Wprowadzenie**

Pomimo wielu prac, których celem jest rozwój sztucznej inteligencji, pozyskiwanie wiedzy eksperckiej z systemów informatycznych wciąż stanowi poważny problem. Technologie Web 2.0 sprawdziły się, jeśli chodzi o prezentację posiadanych zasobów, ale zawiodły przy ich interpretacji. Trafność udostępnianych nam zasobów zależy najczęściej od sprytu człowieka w formułowaniu poszukiwanych haseł. Integracja i interpretacja otrzymanych informacji pozostaje w gestii użytkownika i często nie spełnia jego oczekiwań.

Idea rozwoju sieci semantycznych wciąż pozostaje w fazie projektu, a przeszukiwanie zasobów wiedzy nadal sprawia problem, gdyż wymaga semantycznego rozpoznawania języka.

W artykule przedstawiono analizę niektórych z podejmowanych dotychczas projektów i inicjatyw zmierzających do stworzenia nowej struktury logicznej Internetu, opartej na rozproszonej sieci semantycznej. Zaproponowano również nowe podejście do problemu budowy sieci semantycznej.

---

\* wojciechfolta@gmail.com; a.stolinska@gmail.com.

## 1. Dlaczego potrzebujemy Web 3.0?

Szukając odpowiedzi na to pytanie, warto posłużyć się hipotetyczną sytuacją dotyczącą spontanicznego wyjazdu, na przykład do Pragi. Jak na spontanicznie podjętą decyzję przystało, turystka w godzinę spakowała bagaż i zabrała ze sobą psa. Pierwszy postój nastąpił 50 km przed granicą. Po włączeniu smartfona i wpisaniu w wyszukiwarce pytania: „Jak spędzić 3 wolne dni w Pradze, oczywiście z Dziabągiem?” wyświetliło się kilka linków, a mianowicie: do hotelu w Pradze, w którym mile widziane są zwierzęta, do restauracji z ulubioną kuchnią włoską turystki (oczywiście tylko tych, do których można chodzić z psami), propozycje tras wycieczkowych (jak zwykle uwzględniających zainteresowanie turystki antykwariatami, ale bez muzeów, gdyż nie tolerują w nich obecności czworonogów). „Osobisty internetowy asystent” turystki przewidział, że jej fundusze nie pozwolą tym razem na skorzystanie z hotelu w okolicach centrum (pomógł jej też niedawno zaplanować remont mieszkania). Oto właśnie Web 3.0 – sieć osobista, planująca..., a nawet myśląca? Znająca upodobania, możliwości i uwarunkowania.

Żyjemy w epoce Web 2.0, które w odróżnieniu od nieinteraktywnego Web 1.0 pozwala nam współtworzyć zasoby sieciowe. Nie ma jednoznacznej definicji Web 2.0, nie jest możliwe nawet wskazanie momentu, w którym zaczęła się transformacja. Sieć typu Web 2.0 można jedynie opisać za pomocą kilku cech, takich jak:

- możliwość kreowania treści na stronach internetowych – w postaci blogów czy choćby postów na forach, recenzji zakupionych produktów zamieszczanych za pomocą formularzy;
- rozszerzanie dostępu do Internetu – komputer stracił tę wyłączność na rzecz telefonów komórkowych czy konsoli do gier;
- tworzenie społeczności skupionych wokół serwisów typu Facebook, MySpace czy Nasza-klasa, umożliwiających znajdowanie dawnych znajomych, utrzymywanie kontaktów z ludźmi;
- szybkie i skuteczne sposoby udostępniania treści, na przykład za pomocą serwisu YouTube;
- nowe sposoby pozyskiwania informacji, na przykład za pomocą kanałów RSS<sup>1</sup>.

---

<sup>1</sup> <http://computer.howstuffworks.com/web-30.htm>, 03.06.2011.

Na świecie z Internetu korzysta około dwóch miliardów osób<sup>2</sup> i każda z nich może wzbogacać zasoby sieciowe o niemal dowolną treść. I chociaż nie wszyscy użytkownicy Internetu korzystają w pełni z możliwości, jakie daje technologia Web 2.0, od kilku lat trwają przygotowania do stworzenia nowej generacji sieci WWW, określanej mianem Web 3.0<sup>3</sup>. Nie rozwija się ona w kierunku poszukiwania nowych kanałów komunikacji międzyludzkiej – jej głównym zadaniem jest stworzenie kanału komunikacyjnego między człowiekiem a bazami wiedzy. Interfejsem komunikacyjnym będą przede wszystkim wyszukiwarki, które można porównać do osobistego asystenta: będą znały nasze zainteresowania, nasz potencjał (finansowy i... intelektualny), nasze preferencje. Dzięki temu możliwe będzie szybsze wyszukiwanie informacji, lepsza organizacja czasu i optymalne planowanie.

Historyjka o wyjeździe do Pragi dobrze oddaje ideę Web 3.0. Wyszukiwarki, z których obecnie korzystamy, **nie rozumieją**, czego naprawdę szukamy – znajdują strony na podstawie słów kluczowych, często oferując zestaw zupełnie nieprzydatnych odnośników. Wyszukiwarki oparte na nowej technologii powinny znajdować nie tylko słowa kluczowe, ale również **inteligentnie** zestawiać treści z bazy dostępnych informacji, zgodnie z **indywidualnym profilem internetowym** użytkownika. Nawiązując do przedstawionej powyżej historyjki, można powiedzieć, że wyszukiwarka Web 3.0 będzie wiedziała (pamiętała), że Dziabąg to imię psa, a zatem planując urlop w Pradze, należy uwzględnić specyficzne potrzeby właściciela czworonoga.

Web 3.0 pozwoli zaoszczędzić wiele godzin poświęconych na wyszukiwanie w sieci przydatnych, wartościowych informacji – dokona ich selekcji, zweryfikuje źródła informacji i dopasuje do naszego profilu.

## 2. Sposoby realizacji Web 3.0

Podobnie jak w przypadku sieci generacji Web 2.0, nie ma precyzyjnej definicji Web 3.0, można ją jednak opisać za pomocą cech, wśród których najważniejszą jest chyba wyszukiwanie informacji ze zrozumieniem intencji szukającego oraz kontekstu poszukiwań.

<sup>2</sup> Raport Interactive Bureau Polska (IAB Polska), <http://www.rp.pl/artukul/665029.html>, 13.06.2011.

<sup>3</sup> R. Kurzweil, amerykański naukowiec i futurolog już w 2006 roku ogłosił, że era Web 3.0 rozpocznie się w 2007 roku, <http://www.kurzweilai.net/>, 05. 05.2011.

Z Web 3.0 wiążą się nieodłącznie dwa pojęcia, to jest: *ontologia* i *semantyka*. W informatyce ontologia rozumiana jest jako sposób formalizacji wiedzy; składa się z definicji pojęć (obiektów) i relacji zachodzących między nimi. W uproszczeniu można powiedzieć, że ontologia jest słownikiem, zbiorem powiązanych terminów, na którym dokonuje się kategoryzacji (przyporządkowania obiektów na podstawie ich cech do określonych klas) oraz hierarchizacji (tworzenia struktury klas dziedziczących cechy z klas nadrzędnych). Semantyka w odniesieniu do Web 3.0 rozumiana jest jako dziedzina zajmująca się rozumieniem przez komputery języka naturalnego, co ma bardzo duże znaczenie w organizowaniu baz wiedzy. Dane i semantyczne metadane<sup>4</sup> umożliwiają automatyczne przetwarzanie zawartości znaczeniowej dokumentów.

Analiza literatury i realizowanych projektów pozwala dostrzec trzy ważne kierunki rozwoju Web 3.0, których celem jest rozwiązanie problemu wyszukiwania informacji nie na podstawie słów kluczowych, ale znaczenia wpisanych fraz. Pierwszy i najtrudniejszy w realizacji kierunek związany jest z rozwojem sztucznej inteligencji oraz lingwistyki komputerowej. Dzięki analizie informacji pochodzących na przykład z serwisów społecznościowych i komentarzy maszyni mają znajdować zależności między danymi i wykonywać skomplikowane polecenia i zapytania. Głównym problemem jest w tym przypadku przetwarzanie języka naturalnego (ang. *Natural Language Processing* – NLU). O tym, jak trudny jest to problem, świadczy fakt, że od sześćdziesięciu lat żaden komputer nie przeszedł pozytywnie choćby testu Turinga.

Drugi sposób, który autorzy określają mianem *Web 3.0 social*, wymaga zaangażowania aktywnych użytkowników sieci i twórców, na przykład portali społecznościowych i wyszukiwarek, w odpowiednie opisywanie stron, wykorzystywanie nowych rozwiązań typu protokół Open Graph czy mechanizmy FOAF. To podejście znane jest pod nazwą *microformats* i polega na manualnym rozszerzaniu istniejących stron opartych na (X)HTML o specjalne elementy tego standardu, które uczynią je czytelnymi dla komputerów<sup>5</sup>. Narzędzia te nawiązują do trzeciego kierunku rozwoju Web 3.0, to jest projektu *Semantic Web*, który ma przyczynić się do **utworzenia i rozpowszechnienia standardów** opisywania tre-

<sup>4</sup> Metadane semantyczne opisują znaczenie danych, czyli zawierają informacje o obiektach, postaciach, zdarzeniach.

<sup>5</sup> [http://www.chip.pl/artykuly/porady/2009/06/semantyczna-siec-boty-ucza-sie-kojarzy-c?b\\_start:int=3,07.06.2011](http://www.chip.pl/artykuly/porady/2009/06/semantyczna-siec-boty-ucza-sie-kojarzy-c?b_start:int=3,07.06.2011).

ści w Internecie w taki sposób, by możliwe było ich przetwarzanie przez maszyny i programy z uwzględnieniem znaczenia i kontekstu informacji. Docelowo nowe dokumenty zamieszczane w sieci mają być tworzone zgodnie z określonymi regułami, ale w przyszłości możliwa ma być automatyczna generacja stron WWW, które będą opisane semantycznie. Wiedza konieczna do tego procesu (w którym nie uczestniczy człowiek) będzie czerpana z semantyki zawartej w ontologii.

### 3. Semantyczny opis stron – przykładowe implementacje

Wydaje się, że do stworzenia spersonalizowanej sieci ułatwiającej i przyspieszającej wyszukiwanie informacji wystarczająca byłaby ewolucja istniejącej sieci WWW – przez opisywanie (obudowanie) istniejących zasobów znacznikami oddającymi znaczenie (semantykę) oraz przez klasyfikację i indeksację treści. Proces taki, wykonywany ręcznie lub półautomatycznie, można już obserwować na przykładzie zmian w funkcjonowaniu portalu społecznościowego Facebook i stron z nim powiązanych (dzięki facebookowym wtyczkom) lub projektu DBpedia.

**Protokół Open Graph** przedstawiony został przez firmę Facebook w kwietniu 2010 roku. Jego zadaniem jest umożliwienie twórcom stron zamienienia ich w pewnego rodzaju obiekt Facebooka czy też – bardziej formalnie – obiekt społecznego grafu. Warunkiem koniecznym jest implementacja do danej strony protokołu, czyli dodanie do niej metadanych, którymi przykładowo mogą być: tytuł, pod którym obiekt będzie się pojawiał w grafie (tekst linku), określenie typu obiektu (np. blog czy artykuł), adres URL obiektu i URL obrazka przypisanego do niego, opis strony i inne. Wszystkie informacje metaprotokołu Open Graph mają strukturę: `<meta property="og:{nazwa}" content="{wartość}"/>`<sup>6</sup>.

Protokół Open Graph będąc narzędziem, które umożliwi przekazanie informacji do Facebooka, pozwala na tworzenie połączeń pomiędzy ludźmi. Protokół ten oraz możliwość osadzania facebookowego przycisku „Lubię to” pomagają w specyficznym agregowaniu treści przez łatwiejsze dotarcie do stron polecanych przez znajomych. Jest to niewątpliwie krok w kierunku tworzenia kolejnych elementów sieci semantycznej.

<sup>6</sup> <http://smoku.net/artykuly/wprowadzenie-do-protokolu-open-graph-the-open-graph-protocol>, 07.06. 2011.

**FOAF** (ang. *Friend of a Friend*) to tak zwana lekka ontologia, która umożliwia opis relacji i profili użytkowników serwisów społecznościowych. Jest to pewien standard opisu ludzi, który może być przetwarzany przez maszyny. Zaimplementowane w dokumencie FOAF odwoływanie się do innych tego typu dokumentów (w praktyce do opisów innych osób) sprawia, że możliwe jest tworzenie sieci społecznych.

Odmienne rozwiązania, ale ze wspólnym pierwiastkiem społecznościowym i semantycznym, można znaleźć w projekcie **DBpedia**, którego celem jest lepsze wykorzystanie ogromnych zasobów Wikipedii. Trwa proces ich wyodrębniania, strukturyzowania i dzielenia na klasy – opisywania według odpowiednich ontologii. To „przepisywanie” Wikipedii zmierza do stworzenia bazy wiedzy, na której będą operować „maszyny”, to jest semantyczne aplikacje generacji Web 3.0. Ideę tego mechanizmu dobrze oddaje podtytuł polskiego projektu *DBpedia Polska – społeczna encyklopedia zrozumiała dla maszyn*<sup>7</sup> (rys. 1). Od maja 2011 roku DBpedia Polska umożliwia dostęp do zasobów przez przeglądanie, wyszukiwanie i formułowanie odpowiednich zapytań.



Rys. 1. Strona WWW projektu DBpedia Polska

Źródło: <http://pl.dbpedia.org/>, 12.06.2011.

<sup>7</sup> <http://pl.dbpedia.org/>, 07.06. 2011.

Omawiając w skrócie stan realizacji przekształcania istniejących zasobów w sieć, o której będzie można powiedzieć, że jest bazą wiedzy eksperckiej, nie sposób pominąć projektów, których celem jest stworzenie semantycznych wyszukiwarek. Ich rozwój zmierza w dwóch kierunkach: pierwszy to tworzenie wyszukiwarek analizujących semantycznie strony, zaś celem drugiego jest stworzenie wyszukiwarki przeszukującej zasoby sieci semantycznej.

Przykładami eksperymentalnych wyszukiwarek, których głównym atutem ma być udzielanie odpowiedzi na pytania zadane językiem naturalnym, są Hakia i Powerset. Hakia zwraca odpowiedzi, które są generowane z baz danych, wykorzystując przy tym ontologie budowane na podstawie statystyk uzyskanych w wyniku analizy stron internetowych. Wyszukiwarkami drugiego typu (przeszukują metadane dokumentów) są przykładowo: Swoogle, Sindice i Falcons. Ich znaczenie będzie wzrastać w miarę tego, jak będzie przybywać odpowiednio opisanych stron.

Na małą popularność funkcjonujących wyszukiwarek semantycznych ma wpływ ich wciąż niska skuteczność. Przykładowo polska wyszukiwarka semantyczna KtoCo (<http://ktoco.pl> – od 2009 r. wciąż w wersji testowej), której mechanizm działania opiera się na analizie gramatycznej i semantycznej zadanych pytań z uwzględnieniem polskiej specyfiki językowej, na pytanie: *W jaki sposób działa wyszukiwarka semantyczna?*, zwraca informację: *Znaleziono 0 wyników. Czas wyszukiwania: 5438 ms z listą dwóch (...) paragrafów, które być może zawierają odpowiedź na Twoje pytanie. Należy dodać, że zaproponowane przez wyszukiwarkę strony nie zawierają odpowiedzi na zadane pytanie.*

#### **4. Semantyczny opis zasobów sieciowych – technologie Web 3.0**

Powyżej przedstawiono przykłady implementacji nowych technologii, umożliwiających przetwarzanie metadanych. Zauważono, że metadane semantyczne opisujące zasób można umieścić wewnątrz dokumentu (np. HTML), należy jednak dodać, że mogą być one zamieszczone również w pliku zewnętrznym. W opisanych przykładach realizowanych projektów, zmierzających do stworzenia sieci semantycznej, stosuje się standard RDF (ang. *Resource Description Framework*), uniwersalną specyfikację umożliwiającą uzupełnienie publikowanych treści o odpowiednią semantykę. Każde hasło

podstawowe (podmiot) uzupełniane jest o predykat i jego wartość, tworząc trójkę (ang. *subject, predicate, object*).

Semantyczne znaczenie wyrazów najczęściej osadzone jest w szerokim kontekście, dlatego płaska struktura, jaką tworzy zbiór opisanych powyżej trójek, nie mogła spełnić swego zadania bez uzupełnienia o taksonomię klas i relacji oraz wprowadzanie relacji pomiędzy klasami obiektów i tworzenie w ten sposób języka RDFS (ang. *Resource Description Framework Schema*).

Wieloznaczność dużej liczby wyrazów zawartych w słownikach językowych sprawia, że zdecydowanie łatwiej jest definiować semantykę dokumentów w obrębie poszczególnych dziedzin wiedzy. Realizację tego postulatu wypełnia język OWL (ang. *Web Ontology Language*) – standard konsorcjum W3C.

Forma składniowa języka OWL oparta jest na języku XML, posiada jednak dużą liczbę nowych sposobów zapisu. W języku OWL tworzy się zbiory definicji pojęć, właściwości, obiektów i relacji, czyli tzw. ontologie. Konsekwencją konwencji wielowarstwowości zapisu ontologii w językach jest wielopoziomowa struktura języka OWL:

- najniższy poziom to warstwa OWL Lite, pozwalająca tworzyć prostą taksonomię pojęć ontologicznych oraz relacji z powiązaniem licznymi;
- warstwa OWL DL (ang. *OWL Description Logics*) pozwala na tworzenie rozszerzonych definicji klas z kilkoma rodzajami ograniczeń, które wydatnie wpływają na efektywność i rozstrzygalność systemów opartych na tej warstwie;
- najwyższa warstwa OWL Full jest zbiorem konstruktorów nieposiadających żadnych ograniczeń, co skutkuje brakiem rozstrzygalności.

Kolejną propozycją konsorcjum W3C jest język SWRL (ang. *Semantic Web Rule Language*). Za jego pomocą można reprezentować reguły semantyczne. SWRL rozszerza aksjomaty języka OWL o reguły w postaci klauzul Horna<sup>8</sup>, operujące na pojęciach ontologicznych.

Obecnie ogromnym źródłem wiedzy są bazy danych będące strukturami bardzo dobrze zorganizowanymi, dzięki czemu stosunkowo łatwo można wykorzystać zgromadzone w nich zasoby informacji. Język pozwalający wykorzystać wiedzę z baz danych przez technologie semantyczne to SPARQL (ang. *Simple Protocol And RDF Query Language*). Warto nadmienić, że w projekcie *DBpedia Polska* zapytania formułuje się właśnie w tym języku. Jego składnia

<sup>8</sup> Klauzula Horna to zbiór literalów, z których co najwyżej jeden jest zanegowany.



zbliżona jest do języka SQL. Zapytania zadawane są w postaci grafów RDF uwzględniających zawartą w nich wiedzę, opisaną przez ontologie.

Istotną cechą języka SPARQL jest to, że jest on protokołem sieciowym, dzięki czemu zadawane w nim zapytania mogą być samodzielnie transportowane przez technologie sieciowe.

Technologie semantyczne i semantyczne wyszukiwanie treści rozwijane w Web 3.0 stosowane są w *mashupie* – wirtualnej agregacji informacji pochodzących z wielu źródeł. Pojęcie semantyki w sensie Web 3.0 jest węższe niż semantyka języków naturalnych. Szczegółowość opisu zasobu semantycznego ma pozwolić na wyszukiwanie w nim informacji zgodnie z ich znaczeniem. Semantyka języka naturalnego daje człowiekowi dużo bogatszą wiedzę z wrażeniami estetycznymi włącznie.

Wszystkie proponowane rozwiązania mają wspólną ideę, która polega na rozbudowaniu dotychczasowych zasobów Internetu o dodatkową wiedzę umożliwiającą jej zrozumienie i inteligentne przeszukiwanie. Jednak idea ta może okazać się zgubna. Automatyczne dołączanie nowych przyczynków może stać się furtką, przez którą wprowadzony zostanie współczesny „koń trojański”. W sieci Web 2.0 zamieszczenie dezinformacyjnych stron było stosunkowo łatwe do wykrycia przez użytkowników selekcjonujących i oceniających wartościowość informacji. Należy zauważyć, że w przypadku dzieci i młodzieży korzystających z Internetu tej weryfikacji w zasadzie nie ma. Włączenie do zasobów sieci Web 3.0 stron z nieprawdziwymi informacjami będzie o wiele bardziej brzemienne w skutkach, gdyż dołączone do niej relacje ontologiczne mogą automatycznie lawinowo oddziaływać na tysiące innych prawdziwych zasobów, zmieniając częściowo ich znaczenie, infekując je w nowy i trudny do wykrycia sposób, podważając zaufanie użytkowników do sieci i informacji w niej zamieszczanych.

Rozwiązania wymaga problem zarządzania słownikami ontologicznymi, czyli słownikami hierarchicznymi, taksonomiami z dodanymi relacjami między elementami ontologii lub między ontologiami pozwalającymi opisywać zasoby. Zarządzania tego winny się podjąć odpowiedzialne instytucje, które jednak muszą otworzyć zasoby dla ciągłej aktualizacji dokonywanej przez użytkowników tworzących Web 3.0, z kontrolą wiarygodności wprowadzanej wiedzy.

## 5. Web 3.0 – nowy kierunek, inna perspektywa

Kilkunastoletni rozwój semantycznej sieci WWW prowadzony jest w różnych kierunkach, co historycznie przypomina rozwój wielu innych dziedzin informatyki. Jako przykład można wskazać walkę firm komputerowych o dominację na rynku komputerów osobistych, w której wystartowało wielu konkurentów, a tylko nieliczni osiągnęli sukces czy podobną sytuację związaną z rozwojem systemów operacyjnych. Śledząc przez ostatnie pół wieku powstawanie kolejnych języków programowania, można podjąć próbę symulacji dalszych losów rozwoju sieci semantycznych.

W przytoczonych powyżej przypadkach można było zaobserwować kolejne fazy rozwoju.

Pierwsza z nich to faza poszukiwań, czyli wielokierunkowy rozwój prowadzony przez ośrodki naukowe lub komercyjne firmy, z których każda postawiła sobie za cel uniknięcie błędów swoich poprzedników i stworzenie najlepszego produktu. W tej fazie odpadało zawsze wielu konkurentów, ale nigdy nie wyłoniono jednego zwycięzcy.

Faza druga to faza konsolidacji. Jej przebieg bywał różny, na przykład język Ada miał w założeniach zostać jedynym uniwersalnym językiem programowania. Na realizację tego projektu Departament Obrony USA wydatkował duże środki, a jeszcze większe nakłady finansowe poniosło kilkanaście ośrodków naukowych startujących w konkursie na stworzenie tego języka. W ten sposób powstał produkt na pewno bardzo dobry, który jednak nie został uniwersalnym standardem programistów. Zupełnie inaczej przebiegała walka firm produkujących komputery osobiste, których produkty początkowo były całkowicie niekompatybilne. Tym razem konsolidacja przybrała charakter ostrej walki na kapitalistycznych zasadach o olbrzymi rynek klientów. Prawdziwym kluczem do sukcesu jednej z firm nie było jednak wykupywanie słabszych konkurentów, ale zaskakujące działanie firmy IBM, która nie reagowała na masową produkcję podróbek swoich komputerów, głównie na Dalekim Wschodzie. Donoszono (niesprawdzone pogłoski), że firma IBM celowo ułatwiła dostęp do informacji technicznych o swoich produktach. Zalewanie rynku tanimi podróbkami komputerów osobistych typu IBM pozwoliło zdobyć tej firmie istotną przewagę nad konkurencją, stanowiło bowiem swoistą promocję jej produktów.

Trzecia jest faza monopolistyczna, w której jeden lub kilku zwycięzców „dzieli się łupami”. Można w niej mówić o „budowie na zgłiszczach”, gdyż miliony niezadowolonych klientów bezsilnie usiłują wpłynąć na poprawę nielicznych pozostałych na rynku produktów, jednak działania ich są już wówczas mało skuteczne zarówno w zakresie poprawy jakości, jak i wysokości ceny.

W rozwoju sieci semantycznych również można zaobserwować dwie pierwsze z omówionych faz. Burzliwa faza poszukiwań trwa, a wiele firm i ośrodków rozwija kolejne pomysły. Występują też już próby konsolidacji, i tak rozwijane w ramach projektu Siódmego Programu Ramowego Unii Europejskiej narzędzie LarKC<sup>9</sup> (ang. *Large Knowledge Collider*) ma stworzyć semantyczną sieć WWW, skalowalną do rozmiaru obecnego Internetu. Czy narzędzia powstałe w wyniku tego projektu podzielią los języka Ada, czy powstanie jeden uniwersalny produkt? Wydaje się, że jest to nierealne, gdyż narzędzie to musiałoby być równoważne z pełnym analizatorem języka, jakim posługują się ludzie, a więc najlepiej, żeby działało równocześnie we wszystkich językach naturalnych. W ten sposób zostałby też rozwiązany problem translatorów pomiędzy językami, dzięki uniwersalnemu narzędziu do zrozumienia wiedzy, bez względu na jej leksykalny zapis.

Niestety, pomimo doświadczeń wynikających z niezwykle szybkiego rozwoju wielu dziedzin informatyki, trudno jest sobie wyobrazić, aby w pierwszej połowie XXI wieku sieci semantyczne w fazie konsolidacji osiągnęły poziom realizujący wymienione powyżej postulaty.

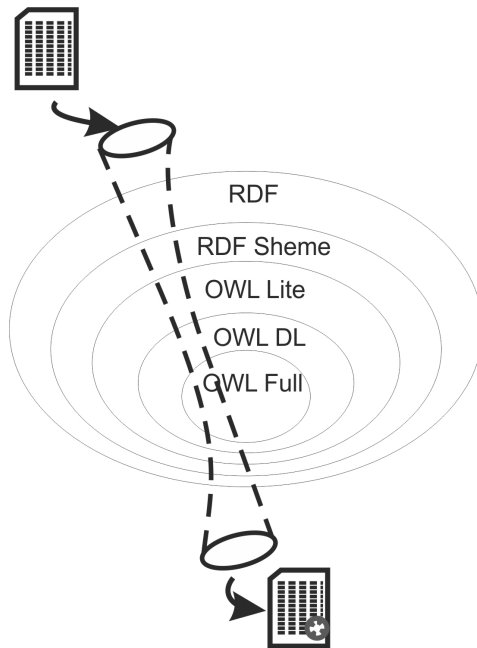
Bardziej realny jest przebieg trzeciej fazy, analogiczny z wprowadzaniem wielu innych rozwiązań komputerowych. Znajdą się potentaci, którzy wykorzystają potrzeby największego obecnie rynku konsumenckiego na świecie i z pozycji monopolisty zaoferują mierny produkt za średnie pieniądze, co przyniesie im krociowe zyski.

Co więc powinna zrobić społeczność internetowa, aby uniknąć naświetlonej powyżej perspektywy? Jeżeli pierwsza część alternatywy, czyli oddanie Web 3.0 w ręce monopolistów nie daje szans na jej swobodny, „zdrowy” rozwój, a druga część oddaje pole hakerom, umożliwiając im dokonanie w krótkim czasie destrukcji sieci, to rozwiązań należy szukać pośrodku.

Nowym podejściem do zbudowania bezpiecznej sieci semantycznej jest zdefiniowanie i zbudowanie agenta – specjalnego programu, którego zadaniem

<sup>9</sup> <http://www.larkc.eu/>, 09.06. 2011.

będzie aksjologiczne wyznaczenie rzeczywistej wartości dokumentu włączanego do zasobów sieci. Ten aksjologiczny agent będzie obliczał parametr prawdziwości dokumentu na podstawie braku sprzeczności z zasobami już zweryfikowanymi (rys. 2). Weryfikacja polegałaby na osiągnięciu progu granicznego parametru prawdziwości i była certyfikatem umożliwiającym wstęp do zasobów globalnych sieci. Podstawową zasadą działania procesu weryfikacji musi być możliwość jego powtarzalności na zlecenie każdego użytkownika sieci, ponieważ parametr prawdziwości jest zależny od zasobów sieci, które ulegają ciągłym zmianom. Postulat ten można zrealizować, powołując instytucje weryfikujące dokumenty, nie jak dotychczas za pomocą udzielanego na stałe (lub określony czas) certyfikatu, lecz dokonując obliczeń sprawdzających prawdziwość dokumentu w chwili otrzymania takiego zlecenia.



Proces obliczenia parametru prawdziwości dokumentu polega na sprawdzeniu, czy nie jest on sprzeczny z dokumentami już zweryfikowanymi i przebiega przez wszystkie warstwy sieci semantycznej. Osiągnięcie progu granicznego parametru prawdziwości jest certyfikatem włączającym dokument do zasobów globalnych sieci.

Rys. 2. Proces obliczania parametru prawdziwości

Źródło: opracowanie własne.

Podstawową zaletą takiego rozwiązania będzie niezależność od instytucji zarządzających. Funkcje zależności wykorzystywane przez aksjologicznego agenta będzie musiała określić odpowiedzialna jednostka, ale standaryzacja problemów, chociażby za pomocą kolejnych dokumentów referencyjnych, jest ogólnie stosowaną praktyką, również w sieciach komputerowych, gdzie liczba corocznie generowanych nowych protokołów nikogo już nie dziwi.

Zaproponowane powyżej rozwiązanie pozwoli na uniknięcie obydwu wymienianych wcześniej zagrożeń. Nie dopuści do zmonopolizowania produktów umożliwiających dostęp do zasobów sieci, broniąc równocześnie przed atakami zagrażającymi integralności nowego żywego organizmu, jakim powinna się stać nowa sieć.

## **Podsumowanie**

W artykule opisano niewielką część spośród wielu inicjatyw podejmowanych w celu stworzenia Web 3.0. Pozwoliło to jednak na zarysowanie kontekstu dla nowego problemu naukowego, który jest przedmiotem badań autorów. Aksjologiczne wartościowanie prawdziwości danych wymaga budowy algorytmów, które bazując na dodatkowych informacjach dołączanych do dokumentów sieciowych, będą prowadzić do obliczenia ich wartości. Może to stanowić nowy dział nie tylko w badaniach podstawowych, ale także w praktycznych badaniach stosowanych.

## **Literatura**

- [http://www.chip.pl/artykuly/porady/2009/06/semantyczna-siec-boty-ucza-sie-kojarzyc?b\\_start:int=3](http://www.chip.pl/artykuly/porady/2009/06/semantyczna-siec-boty-ucza-sie-kojarzyc?b_start:int=3), 07.06.2011.
- <http://computer.howstuffworks.com/web-30.htm>, 03.06.2011.
- <http://www.kurzweilai.net/>, 05.05.2011.
- <http://www.larkc.eu/>, 09.06.2011.
- <http://pl.dbpedia.org/>, 12.06.2011.
- <http://www.rp.pl/artykul/665029.html>, 13.06.2011.
- <http://www.smoku.net/artykuly/wprowadzenie-do-protokolu-open-graph-the-open-graph-protocol>, 07.06.2011.

**CONSTRUCTION OF THE SEMANTIC WEB – A NEW PERSPECTIVE****Summary**

Despite many researches on the development of artificial intelligence, gaining expertise in information systems is still a serious problem. Web 2.0 technologies have succeed in presentation of resources, but failed in their interpretation. From the cunning of man sought in the formulation of terms depends on the accuracy of most available resources to us. Accuracy of resources available to us is dependent on cleverness of a man during formulation of terms of searching. Integration and interpretation of obtained information is user's responsibility and often does not meet his expectations, leaving unsatisfied that the most important information on the subject have failed to even reach.

The idea of the development of Semantic Web is still in phase of project and process of searching knowledge bases still causes a lot of problems, because it requires a recognition of semantic language.

This article presents an analysis of some of the current projects and initiatives aimed at creating a new logical structure of the Internet, based on distributed semantic network. There is also demonstrated a proposal of a new approach to the problem of building the Semantic Web.

**Keywords:** Semantic Web, artificial intelligence, Web 2.0 technologies, Web 3.0 technologies, Web 3.0 social, Open Graph, FOAF, OWL, metadata processing, new ontologies, semantics

*Translated by Wojciech Folta*