

Porównanie z najbardziej podobnymi systemami oraz powiązania z ideami z różnych dziedzin. Przykład z ekonomii

Dodatek do tekstu: „A method for building a scalable centralized ontology - An idea”

Dariusz Daćko

26 kwietnia 2008

www.consensualknowledge.net

1. Porównania

Poniżej opiszę krótko nawiązania do istniejących idei oraz systemów informatycznych, które nie zostały opisane w głównym tekście. Mam nadzieję, że nawiązania te pozwolą łatwiej określić miejsce zaprezentowanej idei wśród istniejących idei.

Filozofia – konsensualna teoria prawdy

(http://en.wikipedia.org/wiki/Consensus_theory_of_truth)

Teoria została wymyślona przez niemieckiego filozofa Jürgena Habermasa (http://en.wikipedia.org/wiki/J%C3%BCrgen_Habermas). Według tej teorii prawdą jest to, odnośnie czego panuje powszechny konsensus wśród ludzi, że jest to prawda. W teorii tej wprowadza się pojęcie „idealnej rozmowy”, w której po dostatecznie długim czasie wszyscy rozmówcy po wymianie wszystkich argumentów powinni dojść do wspólnego stanowiska. Używane są w niej podobne pojęcia oraz zasady do tych z przedstawionej idei:

- 1) pojęcie konsensusu (chyba najważniejsze w tej teorii),
- 2) rozważanie, że cała ludzkość może stanowić grupę osób dyskutujących,
- 3) pojęcie "communicative action" (http://en.wikipedia.org/wiki/Communicative_action).

Ludzie mogą wykonywać różne akcje: komunikować sobie informacje; "reklamować" siebie, idee, produkty; robić coś dla siebie itp. Tylko akcja komunikacji jest całkowicie bezinteresowna. Obecnie w Internecie jest mało bezinteresownych akcji. Większość stron WWW służy ich właścicielom. "Interesowność" szkodzi jakości przekazywanej informacji. Wikipedia z założenia jest bezinteresowna i dlatego jest dobra. Na forach internetowych panuje duża swoboda oraz często dochodzi do wymiany emocji a nie informacji. W zaproponowanym systemie, dzięki zorganizowaniu informacji w ontologię, nie będzie miejsca na wymianę negatywnych emocji. Dyskusja będzie bardziej rzeczowa i konstruktywna.

Filozofowie przeciwni konsensualnej teorii prawdy twierdzą, że konsensus wśród wszystkich ludzi lub pewnej ich podgrupy jest niemożliwy, gdyż jest za dużo ludzi i nie ma technicznych możliwości do przeprowadzenia konstruktywnej dyskusji. Ale według mnie przedstawiony pomysł daje taka możliwość. Nie sądzę żeby filozofowie go rozważali, skoro nawet nie rozważali go informatycy.

Można się łatwo przekonać o poprawności poprzedniego zdania na podstawie omówionego dalej podejścia Open Mind z MIT. W podejściu tym brakuje konsensusu, co skutkuje wandalizmem i dużą pracochłonnością dodawania wiedzy do systemu.

Większość informatyków zajmujących się ontologiami uważa, że nie jest możliwe dojście do konsensusu, gdyż ludzie nie mogą dojść do porozumienia ze względów kulturowych itp. Jednak jeden z wybitnych filozofów jest innego zdania. Oznacza to, że warto sprawdzić poprawność zaprezentowanej idei. Można powiedzieć, że przedstawiona idea jest propozycją komputerowej implementacji konsensualnej teorii prawdy oraz „Sfery publicznej”. Sfera publiczna (http://en.wikipedia.org/wiki/Public_sphere) jest “obszarem życia społecznego, w

którym ludzie mogą zebrać się razem, swobodnie dyskutować, identyfikować problemy społeczne oraz poprzez dyskusję wpływać na działania polityczne”.

Zarządzanie – inteligencja zbiorowa (http://en.wikipedia.org/wiki/Collective_intelligence). Pojęcie opisane w głównym tekście. Przedstawiony pomysł może być potencjalnie interesującym przykładem inteligencji zbiorowej.

Open Space Technology (http://en.wikipedia.org/wiki/Open_Space_Technology) jest sposobem prowadzenia zebrań ludzi umożliwiającym efektywną wymianę poglądów. Można go zastosować na przykład jako alternatywę do konferencji. Podstawową zasadą jest to, że nie ma z góry ustalonego programu zebrania. Porządek zebrania ustalany jest na jego początku i każdy może wnieść do porządku interesujący go problem, krótko go przedstawiając. Następnie uczestnicy zebrania zapisują się na interesujące ich tematy, po czym udają się do osobnych pokoi w celu ich przedyskutowania. Dzięki temu każdy może wypowiedzieć się na interesujący go temat i ewentualnie dojść do wspólnych wniosków.

Także w zaproponowanym systemie każdy zainteresowany będzie mógł wyrazić opinię. Może to się przyczynić do przyspieszenia wymiany wiedzy między ludźmi.

Ontologia Cyc (<http://en.wikipedia.org/wiki/Cyc>) jest największą ontologią zbudowaną do tej pory. Jej wadą jest to, że jest budowana jedynie przez kilkadziesiąt osób (o ile dobrze pamiętam przez 20-30). Do tej pory zapisano w niej w ten sposób około miliona pojedynczych informacji. Jest to ciągle bardzo mało w porównaniu z liczbą informacji zdroworozsądkowej znanej każdemu człowiekowi szacowaną na około pół miliarda pojedynczych informacji. Liczba informacji które mogą być przydatne wszystkim ludziom w tym specjalistom ludziom może sięgać nawet kilku lub kilkadziesiąt miliardów. Oczywiście jest, że zebranie takiej olbrzymiej liczby informacji wymaga zespołowej edycji zgodnie z ideą Web 2.0 (http://en.wikipedia.org/wiki/Web_2). Wikipedia będąca sztandarowym przykładem serwisu Web 2.0 stanowi dobry argument za skalowalnością. Wikipedia w krótkim czasie przerosła swoim rozmiarem Encyklopedię Britannica redagowaną przez wąskie grono specjalistów.

Projekt Citizendium (<http://en.wikipedia.org/wiki/Citizendium>) jest systemem podobnym do Wikipedii realizowanym przez jej współautora. Citizendium ma być alternatywą dla Wikipedii obdarzoną wiarygodnością. Jednym z elementów mających do tego doprowadzić jest uwierzytelnianie użytkowników na podstawie ich danych osobowych. Podobne rozwiązanie jest użyte w Allegro, gdzie jest wymagane uwierzytelnienie na podstawie posiadanego fizycznego adresu pocztowego. Rozwiązanie to ma zlikwidować wandalizm. Analogiczne rozwiązanie zaproponowane jest w przedstawionej idei. Wandalizmu nie udało się uniknąć także w opisanym poniżej systemie Open Mind.

Cechą charakterystyczną Wikipedii jak i Citizendium jest to, że są one encyklopediami. Oznacza to, że wprowadzając do nich informację musimy ją wkomponować w artykuł tak, aby stanowiła całość z innymi informacjami. Powoduje to budowę treści w podejściu top-bottom. Nie można dodać do Wikipedii pojedynczej informacji bez pisania całego artykułu o pewnym pojęciu. Spowalnia to przyrost ilości wiedzy w niej zawartej.

Global brain (http://en.wikipedia.org/wiki/Global_brain) to następna pojęcie przedstawiające podobną ideę do „Sfery publicznej”, jednak tym razem nie od strony polityki i społeczeństwa, a od strony kognitywistycznej oraz Semantycznego Internetu (http://en.wikipedia.org/wiki/Semantic_Web). „Globalny umysł” jest „metaforą inteligentnej sieci stworzonej przez ludzi razem z wiedzą i technologiami komunikacyjnymi łączącymi je”.

Trzy aspekty tego pojęcia zostały opisane na wymienionej stronie w Wikipedii. Jeden z nich, encyklopedyczny, kładzie nacisk na rozwój sieci uniwersalnej wiedzy. Obecnie istniejącym systemem najbardziej zbliżonym do tej wizji jest Wikipedia. Wikipedia jednak jest tylko encyklopedią, czyli zasobem wiedzy ogólnej. Dlatego nie znajdują się w niej nigdy informacje bardzo szczegółowe. Sieć semantyczna (Semantic Web) mogłaby być uważana za „Globalny umysł”, jednak obecnie naukowcy podchodzą do niej sceptycznie z powodu braku pomysłów na jej realizację.

Open Mind (http://en.wikipedia.org/wiki/Open_mind), realizowany przez MIT, jest systemem o celu w pewnym stopniu analogicznym do systemu Cyc. Cyc jest sformalizowaną ontologią. Celem Open Mind jest także utworzenie bazy wiedzy, jednak tym razem mniej sformalizowanej i bardziej zdroworozsądkowej, na przykład typu: „łóżko służy do spania”. W Open Mind zrezygnowano z formalizowania (i tym samym zdolności rozumienia tej wiedzy przez komputer) po to, by wiedza mogła być wprowadzana przez znaczną liczbę internautów. W ten sposób zebrano przeszło milion informacji. Niestety brak sformalizowania oraz brak kontroli użytkowników sprawia, że znaczna część wpisów powtarza się i często dochodzi do wandalizmu. Przykładowo kilkanaście procent użytkowników twierdzi, że „łóżko nie może służyć do spania”, co powoduje, że stwierdzenie „łóżko może służyć do spania” oznaczone jest jako niepewne. Z powodu braku kontroli nad użytkownikami, do uwiarygodnienia informacji potrzebna jest duża liczba głosów nad każdą informacją. Około stu osób musiało zadać sobie trud głosowania nad informacją „łóżko może służyć do spania”, a i tak informacja ta jest niepewna. W zaprezentowanym systemie wystarczyłby jeden głos by dodać wiarygodną informację. Oznacza to, że w tym samym czasie powinno zebrać się kilkadziesiąt razy więcej informacji niż w Open Mind. Dodatkowo korzystanie z takiego systemu byłoby o wiele przyjemniejsze. Wiadomo by było, że nasz wkład jest bardzo ważny, gdyż nie duplikuje się z innymi i nie ma potrzeby jego walidacji. Co więcej system miałby praktyczne zastosowanie, a więc wiele ludzi miało by cel w jego rozwoju. Można z tego wyciągnąć wniosek, że w czasie potrzebnym na rozwój systemu Open Mind, do zaproponowanego systemu udałoby się wprowadzić zasób wiedzy zdroworozsądkowej porównywanej z zasobem wiedzy człowieka. Wiedza specjalistyczna, nieznaną laikom jest bardziej sformalizowana i mniej rozmyta, a więc jej zbieranie byłoby łatwiejsze. System byłby wiarygodnym źródłem informacji, dlatego naukowcy chętnie by z niego korzystali w przeciwieństwie do obecnej Wikipedii i Open Mind. Daje to szansę na to, że w systemie mogłaby zostać zapisana także wiedza specjalistyczna, taka jak na przykład właściwości poszczególnych algorytmów sortowania, właściwości poszczególnych programów do porównywania plików, czy też wiedza matematyczna lub astronomiczna.

Mindpixel (<http://en.wikipedia.org/wiki/Mindpixel>) był podobną inicjatywą do Open Mind. W przeciwieństwie do Open Mind posiadał pewną kontrolę użytkowników. Jednak kontrola wiązała się z kosztem, gdyż każda z informacji musiała być wprowadzona przynajmniej przez dwudziestu użytkowników.

Podsumowując istnieją różne idee budowy dużej bazy wiedzy oraz kilka prób ich realizacji. Próbowano utworzyć bazę wiedzy różnymi metodami wykorzystującymi różne zasady. Wydaje się, że warto byłoby ją zbudować wykorzystując najlepsze zasady użyte w poszczególnych systemach. Z systemu Cyc można by wykorzystać ustrukturalizowanie wiedzy tak, aby wprowadzana wiedza nie powtarzała się w systemie, a jej odnalezienie i wykorzystanie było łatwe. Z projektu Citizendum można wykorzystać zasadę reprezentowania użytkowników przez realne osoby. Dzięki temu uzyskano by pewność przechowywanej informacji bez związanego z tym kosztu takiego jak w Mindpixel. Jak

można przeczytać w głównym tekście prawdopodobnie nie spowoduje to odgadnięcia autora informacji. Możliwe, że w przypadku włamania do systemu można by było uniknąć także odnalezienia wszystkich wpisów danej osoby dzięki użyciu technik kryptograficznych.

Wydaje się, że uniknięcie nadmiarowości wprowadzania informacji, pewność wprowadzonej informacji oraz szybkość jej wprowadzania dzięki wiarygodności użytkowników mogłyby umożliwić zbudowanie systemu przechowującego o wiele więcej informacji niż obecnie istniejące systemy.

2. Zastosowanie w ekonomii

W głównym tekście zostały opisane implikacje dla ekonomii wynikające szczególnie z semantycznego opisu produktów. Poniżej zostanie dodatkowo opisany jeszcze jeden aspekt związany z automatycznym uzyskiwaniem oceny produktu.

Każdy produkt opisany jest numerem EAN składającym się z numeru kraju (np. 590 dla Polski), numeru producenta (np. 0500 dla Hortexu), pięciocyfrowego numeru produktu danego producenta (np. 02427 dla 1-litrowego nektaru czarna porzeczka Hortexu) oraz jednej cyfry kontrolnej. Przykładowe kody to:

Nektar Hortex Czarna porzeczka 1l	590 0500 02427 6
Nektar Hortex Greipfrut rubinowy 1l	590 0500 02431 3
Nektar Hortex Multiwitamina 1l	590 0500 02432 0
Nektar Hortex Wielowocowy 1l	590 0500 02438 2
Nektar Hortex Brzoskwiniowy 1l	590 0500 02440 5
Nektar Hortex Bananowy 1l	590 0500 02441 2

Cechy nektaru z czarnych porzeczek podane na jego opakowaniu to między innymi skład:

- sok a czarnych porzeczek (25%),
- woda,
- cukier,
- naturalny aromat z czarnych porzeczek.

Jeśli produkt ma w nazwie wyraz nektar lub sok, to mamy prawnie zagwarantowaną pewność, że w rzeczywistości zawiera naturalny sok a nie związki chemiczne nienaturalnego pochodzenia. Zastanawiać nas może składnik „naturalny aromat z czarnych porzeczek”. Dlaczego nie wystarczy sam sok, tylko dodatkowo jest on jeszcze aromatyzowany? W jaki sposób ten aromat jest otrzymywany? W zaproponowanym systemie będzie można się tego dowiedzieć. W polu składniki pojęcia „Nektar Hortex Czarna porzeczka” wybierzemy składnik czwarty i sprawdzimy jego znaczenie. Oczywiście w tym przypadku taka dokładność wydaje się zbyt dużą przesadą, jednak na przykład co w przypadku wędlin? Ile konserwantów i barwników zawiera wędlina wybrana przez nas na podstawie jej ładnego wyglądu? Jeśli dużo, to może wybrać taką wędlinę, która ich nie zawiera? Może inny producent nie używa tylu konserwantów i barwników? Tylko który? Powracając do przykładu nektaru może się okazać, że nektar sam w sobie jest zdrowy, ale jego opakowanie nie jest biodegradowalne lub producent używa w czasie produkcji nieekologicznych technologii.

Kod EAN ułatwi odnalezienie interesującej nas informacji. Zamiast szukać w przedstawionym systemie obiektu reprezentującego interesujący nas produkt, wystarczy poszukać go po kodzie EAN. Można sobie także wyobrazić moduł Bluetooth lub oprogramowanie do kamer internetowych automatycznie odczytujące kod kreskowy z opakowania produktu i przesyłające kod EAN do telefonu komórkowego (wyposażonego w łącze z Internetem) lub komputera. Zamiast wyświetlać wszystkie cechy produktu można by wyświetlić zbiorczą informację o tym czy produkt jest: a) zdrowy, b) jego produkcja jest ekologiczna, c) spełnia inne dowolne kryteria ustalone przez nas wcześniej itd. Ze względu na użycie ontologii określenie kryteriów będzie łatwe. Wystarczy że konsument określi, że chce kupować „zdrową żywność”, a na podstawie ontologii zostanie automatycznie podjęta

decyzja co jest zdrowe. Przykładowo konserwant Ennn jest zdrowy a Emmm nie jest, a przynajmniej jest mniej zdrowy niż poprzedni. W przypadku nie spełniania naszych kryteriów mogłoby to być sygnalizowane przez czerwony wskaźnik na module Bluetooth, telefonie komórkowym lub w komputerze. Równocześnie mógłby zostać zaproponowany lepszy produkt. W dobie rozwoju przemysłu oraz marketingu rozwiązanie takie może znacząco wpłynąć na to, że producenci będą wytwarzać lepsze, zdrowsze i bardziej ekologiczne produkty, a konsumenci będą kupować bardziej świadomie i od lepszych producentów. Może to wpłynąć zarówno na zdrowie ludzi jak i ochronę środowiska. Ideę można zastosować do produktów ze wszystkich branż: spożywczej, elektroniki itd.