

Semantyczny marketing - baza wiedzy o produktach

Dariusz Daćko

26 października 2008

zmiany 12 sierpnia 2010

www.consensualknowledge.net

1. Zakup produktu

W 2001 roku chciałem kupić drukarkę atramentową do zastosowań domowych. Po przeczytaniu rankingu w jednym z czasopism komputerowych wybrałem jeden z modeli. Mój wybór wynikał z tego, że drukarka była dość znanej firmy oraz według artykułu drukarka szybko i tanio drukowała, a wydruk był dobrej jakości. Głowice były częścią drukarki a nie kartridża z tuszem. Dzięki temu dokupowany tusz był tańszy i tym samym druk był ekonomiczny. Dane zawarte w artykule potwierdziłem w ulotce reklamowej producenta. Opiswany model drukarki okazał się najlepszy, więc kupiłem go. Po przeczytaniu instrukcji obsługi dowiedziałem się, że drukarka rzeczywiście drukuje 15 stron na minutę, ale dopiero po zakupieniu dodatkowej głowicy. Niestety głowica kosztowała prawie tyle samo, co sama drukarka. Oczywiście zakup tej głowicy nie wchodził w grę, gdyż zależało mi na oszczędnym drukowaniu. Tym sposobem ja oraz redakcja czasopisma zostaliśmy wprowadzeni w błąd, a ja dodatkowo stałem się posiadaczem drukarki drukującej dwie czarno-białe strony na minutę. Drukowanie 120-stronicowych materiałów trwało godzinę. Ponieważ wolałem czytać materiały na papierze niż na monitorze, często musiałem czekać na ich wydrukowanie. Pewnego dnia wyjechałem na dwumiesięczne wakacje. Po powrocie okazało się, że dysze drukarki zaschły. Niestety nie udało się ich odetkać. W serwisie dowiedziałem się, że aby dysze nie zasychały powinno się od czasu do czasu drukować. Nie byłem tego świadomy, gdyż nie napisano o tym w instrukcji obsługi. Mimo to dwumiesięczny okres nie powinien nieodwracalnie zatkać dysz. Niestety dowiedziałem się, że mój model drukarek był szczególnie podatny na zapychanie dysz.

Jestem przekonany o tym, że producent działał świadomie podając informację o szybkim drukowaniu oraz pomijając informację o tym, że można to osiągnąć jedynie kupując bardzo drogą głowicę. Jest to oczywiste, że postąpił tak by nie zrazić klienta do zakupu. Tym samym producent dbał o interesy swoje a nie konsumenta.

2. Potrzebna informacja i dostępne źródła

Proces wyboru odpowiedniego produktu jest trudny. Produkt powinien być jak najlepszy dla nas, a jednocześnie nie szkodzić środowisku. Powinien być użyteczny, zdrowy (na przykład produkty spożywcze nie powinny zawierać szkodliwych konserwantów) i tani. Sam produkt jak i jego opakowanie powinno być biodegradowalne, a proces jego produkcji powinien być ekologiczny. Możemy także chcieć wspierać na przykład lokalnych producentów. Musimy dokonać wielokryterialnej decyzji na podstawie dużej liczby niepewnych oraz często ze sobą sprzecznych lub powtarzających się informacji. Na temat niektórych cech produktów możemy w ogóle nie mieć informacji. Materiały reklamowe najczęściej zawierają jedynie informację sprzyjającą producentowi lub sprzedawcy. Na przykład w ulotce dla odkurzacza zawierającego regulację mocy znajdziemy informację o tym, że posiada on taką funkcję. Jednak jeśli odkurzacz nie ma regulatora mocy, to najczęściej nie zostanie to jawnie napisane. Ponieważ taka informacja jest kopiowana do wielu stron WWW, mamy do czynienia z dużą nadmiarowością. Dlatego najczęściej

znajdujemy opisy produktu nie zawierające szukanej przez nas informacji, która świadczyłaby o tym, że produkt nie jest najlepszy. Podsumowując należy ocenić produkt pod wieloma kryteriami, jednocześnie mając bardzo utrudniony dostęp do informacji. Najczęściej powoduje to podjęcie nienajlepszej decyzji.

W dalszej części zostanie opisany projekt systemu komputerowego umożliwiającego konsumentowi dotarcie do obiektywnej i pełnej informacji o poszczególnych produktach. Dzięki temu klient będzie mógł podejmować bardziej świadome decyzje i kupować najlepsze dla niego produkty. Z kolei zmiany w popycie powinny zwiększyć jakość produkowanych towarów przyczyniając się do pozytywnych zmian na całym rynku.

3. Web 2.0 jest lepszym źródłem informacji

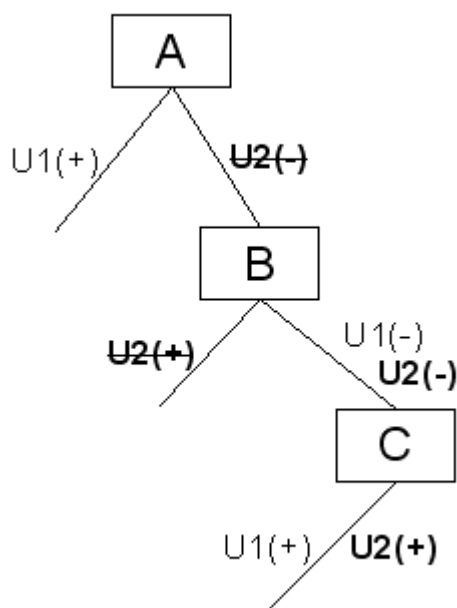
Większość produktów to pojęcia o mniejszej różnorodności niż pozostałe pojęcia języka naturalnego. Na przykład drzewa określonej odmiany mogą mieć różnie ukształtowane korony oraz różną liczbę konarów. Produkty przeważnie mają tę samą charakterystykę zaraz po ich wyprodukowaniu. Odkurzacze pewnego typu mogą mieć pewną cechę (na przykład regulację mocy) lub jej nie mieć, ale wszystkie odkurzacze tego typu są takie same. Oczywiście używane odkurzacze mogą się różnić tym, że w jednym egzemplarzu mogło coś się zepsuć. Jednak mimo to egzemplarze danego modelu są mało zróżnicowane. Dzięki temu każdy produkt ma ściśle określoną charakterystykę. Informacja o produktach nie jest dwuznaczna. Wystarczy tylko do niej dotrzeć i ją zapisać, a będzie ona dostępna dla wszystkich konsumentów. Jak już wcześniej pokazano producenci oraz sprzedawcy nie są dobrym źródłem informacji o produktach. O wiele lepszym źródłem są konsumenci.

Od niedawna w Internecie panuje trend Web 2.0 polegający na aktywnym tworzeniu treści Internetu przez internautów. Można by utworzyć serwis Web 2.0 umożliwiający internautom dodawanie informacji o produktach. Dzięki temu powstałaby bardzo duża, obiektywna baza informacji o wszystkich produktach. Każdy użytkownik miałby możliwość dodawania informacji.

4. Określanie prawdziwości informacji

Jak wcześniej pokazano większość informacji o produktach można zweryfikować i określić ich poprawność. Jest to ważne, gdyż w przeciwieństwie do producenta internauci mogą się częściej mylić. Niektórzy z nich poświęcą mniej czasu na sprawdzenie dodawanej informacji i dlatego może ona być błędna. Często się też zdarza, że użytkownicy świadomie dodają błędne informacje (wandalizm). To że internauci wprowadzają do Internetu informacje o różnej jakości widać najlepiej w obecnie istniejącym Internecie.

Jeśli jeden z użytkowników (oznaczymy go jako U1) dodałby do systemu pewną informację (oznaczymy ją jako informację A, patrz rysunek 1), to każdy komu wydaje się że jest ona błędna (np. użytkownik U2) może zgłosić za niepoprawnością tej informacji. Jednak dodatkowo musi podać argument przeciwko poprawności informacji (np. argument B). Pierwotny autor informacji (U1) oraz wszyscy użytkownicy którzy oznaczyli informację A jako poprawną muszą odnieść się do podanego przed chwilą argumentu B. Każdy argument jest zwykłą informacją. Dodatkowo jest on połączony z pierwotną informacją (A) jako argument przeciwko niej. Użytkownik U1 optujący za poprawnością informacji A powinien dowieść fałszywości informacji argumentu (B) lub tego, że poprawny argument B wcale nie świadczy o poprawności informacji A. Jeśli wybierze pierwszą możliwość, to musi podać argument przeciwko informacji B (np. argument C). Użytkownik U2 po zapoznaniu się z argumentem C może przyznać rację użytkownikowi U1 i oznaczyć informację B jako nieprawdziwą. Dzięki temu wszyscy użytkownicy systemu zainteresowani informacjami A i B dochodzą do konsensusu. Według nich, informacja A ostatecznie okazała się poprawna, a informacja B niepoprawna. Konsensus osiągnięto dzięki dyskusji na argumenty.



Rysunek 1. Drzewo argumentacji. Informacje dodane do systemu przedstawione są jako prostokąty. Gałąź wychodząca od prostokąta w dół oznacza, że jeden z użytkowników optuje za lub przeciw tej informacji. Identyfikator użytkownika oraz to czy głosuje za czy przeciw pokazane jest dzięki etykietce z boku gałęzi. Pierwszy głos za poprawnością informacji może mieć argument ale nie musi. Argumenty traktowane są jak zwyczajne informacje. Jeśli gałąź ma kilka etykiet, to znaczy to, że kilku użytkowników oddało swój głos popierając go tym samym argumentem. Użytkownik który doszedł do wniosku, że dodana przez niego informacja jest błędna może zmienić wartość głosowania usuwając swój głos z relacji „za” poprawnością (na rysunku przekreślona etykieta) i ew. dodając głos do relacji „przeciw”.

Użytkownicy będą mieli tym większą wiarygodność im więcej wprowadzili poprawnych informacji oraz im mniej wprowadzili informacji, które okazały się niepoprawne. Jeśli część użytkowników optuje za poprawnością informacji a część przeciw, to informacja ta nie ma ustalonej poprawności. W takim przypadku będzie obliczane prawdopodobieństwo poprawności uwzględniając wiarygodność użytkowników optujących za i przeciw. Dzięki temu każdy głos użytkownika będzie w systemie odpowiednio uwzględniony.

Jeśli pewnemu użytkownikowi udowodni się wandalizm, czyli celowe wprowadzanie w błąd, to będzie go można na dłuższy okres (na przykład na dwa lata) usunąć z systemu. Aby uniknąć powtórnego zarejestrowania się pod innym loginem powinno się weryfikować użytkowników na przykład na podstawie miejsca zamieszkania, tak jak to ma miejsce w przypadku Allegro.

Powróćmy na chwilę do przykładu omówionego na początku tekstu. Załóżmy, że producent dodał informację o tym, że jego drukarka kosztująca określoną ilość pieniędzy drukuje 20 stron na minutę. Po zakupie drukarki i przeczytaniu instrukcji obsługi jeden z klientów oceni tą informację jako fałszywą podając fragment z instrukcji obsługi mówiący, że standardowe wyposażenie drukarki (w cenie zapisanej w systemie przez producenta) umożliwia drukowanie jedynie 2 stron na minutę. Producent będzie zmuszony przyznać się do błędu i poprawić błędną informację. Jeśli tego nie zrobi, internauci ocenią tą informację samodzielnie wieloma głosami jako nieprawdziwą. W ten sposób producent straci na wiarygodności, gdyż wprowadził błędną informację. Dodatkowo konsumenci nie będą już wprowadzani w błąd. Oczywiście producent mając świadomość konsekwencji prawdopodobnie nie zdecyduje się na dodanie błędnej informacji do systemu.

Większość informacji o produktach można w naturalny sposób przedstawić jako wartości ich cech. Na przykład waga produktu jest przedstawiana za pomocą liczby. Dzięki temu, że informacja jest drobnoziarnista można łatwiej określić jej prawdziwość. Określenie prawdziwości komentarza składającego się z kilku zdań byłoby trudniejsze, gdyż komentarz mógłby częściowo zawierać informację poprawną a częściowo błędną. Ocena nie mogłaby być wyrażona jako prawda lub fałsz. Dlatego wydaje się, że drobnoziarnistość jest kluczową

cechą umożliwiającą dojście do konsensusu. Jest to też cecha różniąca proponowany system od istniejących systemów umożliwiających dodawanie przez internautów komentarzy na temat wybranych produktów.

5. Cechy bazy

W systemie może być bardzo dużo informacji. W celu ułatwienia ich wyszukiwania powinno się je uporządkować w hierarchiczny system organizacji wiedzy zwany siecią semantyczną. W systemie takim definiuje się obiekty, ich właściwości oraz relacje między obiektami, w tym relacje dziedziczenia. Na przykład można zdefiniować trzy pojęcia: „Drukarka”, „Drukarka atramentowa” oraz „LG model A”. Między tymi pojęciami występują relacje dziedziczenia. Dla pojęcia „Drukarka” definiujemy między innymi cechę „liczba stron czarno-białych na minutę”. Dla pojęcia „Drukarka atramentowa” nie musimy już definiować tej cechy. Zdefiniujemy za to na przykład cechę „zużycie atramentu na 100 stron” albo „liczba kolorów atramentu”. Dzięki temu dodając do systemu pojęcie „LG model A” prawdopodobnie nie musimy już definiować żadnych nowych cech – wystarczy wypełnić tylko brakujące dane nie odziedziczone od pojęć nadrzędnych. Dzięki hierarchicznej organizacji pojęć będzie można szybko odnaleźć szukaną informację, a dodawane informacje nie będą się duplikowały. W przypadku gdy dodawaną informację trudno przedstawić jako wartość cechy, można ją dodać jako zdanie w języku naturalnym. Jednak należy je powiązać z obiektami których dotyczy w celu ułatwienia wyszukiwania.

Organizacja informacji ma jeszcze jedną dodatkową zaletę. Jeśli chcemy dowiedzieć się wszystkiego o interesującym nas produkcie, wystarczy zlokalizować go w hierarchii. Wszystkie informacje zapisane w systemie na ten temat będą powiązane ze znalezionym obiektem. Już nie trzeba będzie mozolnie przeszukiwać nadmiarowe strony w Internecie. Jeśli nie znajdziemy informacji na stronie danego pojęcia, to oznacza to, że informacji tej w systemie nie ma – nie trzeba szukać w innych miejscach systemu.

Użytkownicy będą mieli określoną wiarygodność oraz będą o nią dbali. Nawet jeśli tylko jeden użytkownik dodał pewną informację, to można uznać ją za wiarygodną pod warunkiem, że użytkownik jest wiarygodny. Informacja nie musi być sprawdzana przez nikogo, żeby stała się wiarygodna. Oznacza to, że przyrost informacji w systemie może być bardzo szybki. Dzięki temu można też dodawać regionalne produkty, a nie tylko te dostępne na rynku globalnym.

6. Zastosowania bazy

Dotychczas wspomniano o jednym zastosowaniu bazy wiedzy o produktach: bazę można wykorzystać do odnalezienia konkretnej poszukiwanej informacji. Po pewnym czasie można się spodziewać, że wiedza odnośnie najpopularniejszych produktów będzie kompletna, to znaczy prawie wszystkie cechy będą miały określone wartości. Ponadto wiedza będzie pewna, czyli w bazie nie będzie niepoprawnych informacji. Wiedzę o tak wysokiej jakości można wykorzystać do automatycznego określania najlepszego produktu spełniającego preferencje użytkownika. Już nie trzeba będzie samodzielnie porównywać wartości cech wszystkich produktów. Wystarczy, że użytkownik poda swoje preferencje, a system sam znajdzie najlepszy dla użytkownika produkt. Aby skorzystać z automatycznego wyszukiwania produktów po pierwsze powinno się uzupełnić w profilu niezmiennie preferencje. Na przykład można podać następujące ogólne informacje:

- użytkownik nie jest zamożny (należy podać sumę określającą, co to mniej więcej oznacza),
- użytkownikowi zależy na ochronie środowiska,
- miejsce zamieszkania.

Dodatkowo podczas szukania drukarki można podać poniższe informacje:

- użytkownik szuka drukarki do zastosowań domowych,

- użytkownik będzie drukować dużo stron czarno-białych, a więc zależy mu na szybkim drukowaniu takich stron.

Odpowiednia wtyczka do systemu posiadająca wiedzę dziedzinową (w tym przypadku na temat drukarek) weźmie pod uwagę preferencje użytkownika i wybierze najlepsze produkty. Z faktu, że użytkownik nie jest zamożny wynika, że cena jest kryterium posiadającym dużą wagę. Z faktu, że zależy mu na ochronie środowiska wynika, że sam produkt oraz jego opakowanie powinny być biodegradowalne, a proces produkcji powinien być ekologiczny. Oczywiście drukarka powinna być ogólnie dobra (jakość, szybkość itd.) – te wszystkie reguły będą zawarte we wtyczce. Dodatkowo wtyczka zwróci szczególną uwagę na szybkość druku. Produkty będące rezultatem poszukiwań mogą być przedstawione do porównania na wynikowej stronie. Można tam też objaśnić znaczenia poszczególnych cech produktów oraz skutki wynikające z różnic między nimi. Na wynikowej stronie będzie też można od razu kupić wybrany produkt. Wtyczki mogą cechować się dowolną złożonością. Najmniej skomplikowane będą uwzględniały jedynie podstawowe reguły. Najbardziej złożone mogą zbliżać się wiedzą do eksperta w danej dziedzinie. Będzie tu duże pole do popisu dla programistów.

Bazę będzie można wykorzystać także do wyświetlenia informacji o sprzedawanym produkcie na stronach sklepu, serwisu aukcyjnego itp. Pracownicy sklepu albo sprzedający na aukcji nie będą musieli samodzielnie redagować opisu produktu albo szukać odpowiedniego opisu w sieci WWW. Wystarczy, że podadzą identyfikator produktu w bazie wiedzy, a dane zostaną wyświetlone na stronie.

Pewnym ułatwieniem dla kupujących na aukcjach może być wprowadzenie dodatkowego pojęcia „używanego produktu”. Jeśli produkt jest używany, to może znacząco różnić się od nieużywanego. Cechy używanych produktów mogą być wspólne dla wszystkich produktów na przykład: stopień zużycia, jak długo produkt był używany, czy sprzedający jest pierwszym właścicielem itp. Cechy mogą być także specyficzne dla wybranego typu produktów. Na przykład dla samochodów będzie to: przebieg, informacja czy samochód jest po wypadku; a dla telewizora: zużycie luminoforu ekranu. Jeśli sprzedający nie wypełni pewnego pola, to będzie to sygnał informujący kupującego o tym, że sprzedający mógł chcieć zataić tę informację. W ten sposób uniknie się także niejasnych sformułowań kamuflujących złe cechy sprzedawanego egzemplarza produktu.

System może być wykorzystywany także w urządzeniach mobilnych. Załóżmy, że posiadamy miniaturowy czytnik kodów kreskowych wyposażony w interfejs Bluetooth oraz telefon komórkowy z łączem do Internetu. Czytnik kodów mógłby odczytać kod interesującego nas produktu i przesłać go do telefonu komórkowego. Opcjonalnie czynność tą mógłby samodzielnie wykonać telefon komórkowy wyposażony w aparat fotograficzny. Następnie telefon komórkowy sprawdziłby czy interesujący nas produkt spełnia kryteria podane wcześniej przez nas w profilu. W przypadku nie spełnienia podanych kryteriów mogłoby to być sygnalizowane przez czerwony wskaźnik na module Bluetooth lub w telefonie komórkowym. Równocześnie mógłby zostać zaproponowany lepszy produkt. Spełnienie kryteriów mogłoby być sygnalizowane zielonym wskaźnikiem. W takim przypadku moglibyśmy włożyć go do koszyka. Oczywiście korzystalibyśmy z bazy jedynie przy pierwszym zakupie nieznanego nam produktu.

Do systemu zostaną dodane także informacje o tym, jakie produkty są dostępne w jakich sklepach. Jeśli sklep nie posiada systemu zarządzania stanem produktów, to może dodać do bazy jedynie kategorie dostępnych w nim produktów. Mogą być dodane także informacje o godzinach otwarcia poszczególnych sklepów. Dzięki temu system mobilny może przekazać nam informację do którego sklepu znajdującego się w pobliżu nas powinniśmy się udać aby najtaniej kupić wybrany produkt lub gdzie w pobliżu nas znajduje się tania apteka czynna w sobotę o godzinie 20. Obecnie informacja o poszczególnych sklepach jest często

rozproszona po wielu stronach WWW w celu zwiększenia szansy odnalezienia jej przez potencjalnego klienta. Ale co się stanie jeśli jeden ze sklepów zostanie zlikwidowany albo godziny otwarcia zostaną zmienione. Nikt nie będzie uaktualniał dziesiątków wpisów. Aktualizacja informacji o dowolnym sklepie będzie prosta w scentralizowanej bazie produktów. Jeśli zmieni się godzina zamknięcia sklepu, to trzeba będzie uaktualnić tylko jedną liczbę, a wszyscy będą widzieć aktualną informację.

7. Skutki

Jak już wcześniej wspomniano zbudowanie bazy wiedzy o produktach może mieć zarówno lokalny jak i globalny skutek. Lokalnie może ona przynieść korzyść konsumentowi, który dzięki niej wybierze lepszy i bardziej dla niego odpowiedni produkt. Popyt rządzi podażą. Dlatego jeśli baza będzie używana przez wiele osób, to może to spowodować globalne skutki ekonomiczne na całym rynku. Producenci będą dopasowywać się do potrzeb konsumentów wytwarzając lepsze, zdrowsze i bardziej ekologiczne produkty. Konkurencja w zaspokajaniu tych potrzeb będzie miała pozytywny wpływ na kierunek dalszego rozwoju gospodarki, zdrowie ludzi oraz ochronę środowiska. Dobrobyt ludzi powinien zwiększyć się dzięki efektywniejszemu zaspokajaniu mniejszym kosztem naszych potrzeb.

Oprócz opisanych korzyści baza może przynieść dodatkową korzyść jej realizatorowi. Po wyborze odpowiedniego produktu klient ma możliwość złożenia zamówienia. Ze względu na globalny charakter zyski ze sprzedaży zamawianych produktów mogą być dość znaczne.

8. Ogólnotematyczna baza wiedzy

Powyżej opisana idea jest zawężeniem idei budowy ogólnotematycznej bazy wiedzy do pojęć opisujących produkty. Pierwszą z tych baz opisano w tekście [1]. Poza wymienioną różnicą pomysły są takie same. Dlatego w celu dokładniejszego zapoznania się z metodą zarządzania wiedzą można przeczytać ogólniejszy tekst. Idea została opisana na nowo z punktu widzenia produktów, gdyż:

- 1) Jak już napisano cechy produktów mają ściśle określone wartości. Dlatego łatwiej będzie przekonać do tego, że konsensus odnośnie zawartości bazy jest możliwy.
- 2) Informacja o produktach w naturalny sposób podlega strukturalizacji. Cechy produktów często przedstawia się w postaci wyliczenia cecha-wartość. Ze względu na jasno określone cechy produktów łatwo określić relacje hierarchiczne między różnymi klasami produktów.
- 3) Budowa bazy wiedzy o produktach jest ekonomicznie uzasadniona.
- 4) Projekt jest samofinansujący się, gdyż koszty poniesione podczas jego realizacji zwrócą się w trakcie jego używania.

Doświadczenia uzyskane podczas realizacji bazy wiedzy o produktach mogą być wykorzystane do zrealizowania ogólnej bazy wiedzy, która posiada wiele dodatkowych zalet i zastosowań w stosunku do tej pierwszej.

9. Przykład

Poniżej zostaną przedstawione przykładowe pojęcia dotyczące odkurzacza. Uproszczone strukturę pojęć, gdyż prezentowanie pełnej struktury nie było konieczne w celu przekazania idei. Do bazy produktów dodano następujące cztery pojęcia:

- a) pojęcie „Produkt” o następujących zdefiniowanych cechach:
 - seria,
 - kolor o typie „Kolor”,
 - producent o typie „Przedsiębiorstwo”,
- b) pojęcie „Produkt elektryczny” będące hiponimem (tj. pojęciem dziedziczącym) pojęcia „Produkt” o następujących cechach dodatkowych:

- moc (o jednostce Wat),
- c) pojęcie „Odkurzacz” będące hiponimem pojęcia „Produkt” o następujących cechach dodatkowych:
 - „ma filtr HEPA” o typie logicznym,
 - „ma regulację mocy” o typie logicznym,
- d) pojęcie „V-CP753NBB” będące hiponimem pojęć: „Produkt elektryczny” i „Odkurzacz” o następujących wartościach wcześniej zdefiniowanych cech:
 - seria: Turbo Zet (pojęcie)
 - producent: LG (pojęcie)
 - „ma filtr HEPA”: nie
 - „ma regulację mocy”: nie
 - moc: 1500 W
 - kolor: niebieski (pojęcie)

Po wyborze pojęcia „Turbo Zet” mogą być skopiowane wartości wspólne dla całej serii, tj. informacja o producencie, filtry HEPA i regulacji mocy. Jeśli wartość jest pojęciem, to oznacza to, że jest wybierana z listy zdefiniowanych pojęć danego typu.

Literatura

1. Dacko D.: *A method for building a scalable centralized ontology*. Dostępny pod adresem: www.consensualknowledge.net/files/centralized_ontology.pdf.

Zmiany w tekście

- 12 sierpnia 2010 – „ontologią” zamienione na „siecią semantyczną”