

**Kinga Kajfosz**

Uniwersytet Ekonomiczny w Katowicach

---

## ZARYS ONTOLOGII DO OCENY WIARYGODNOŚCI INFORMACJI POZYSKANEJ Z INTERNETU

---

**Streszczenie:** W artykule poruszono ważną kwestię wiarygodności informacji pozyskanej z sieci. Ukazano w nim trudności występujące podczas oceny wiarygodności informacji oraz zaprezentowano opracowane metody oceny. W artykule przedstawiono także dwie ontologie mające na celu weryfikację informacji. W pierwszej z nich, utworzonej przez Foxa i Huanga skoncentrowano się na pochodzeniu wiedzy, na sprawdzeniu autora informacji oraz jej powielania w innych źródłach. Uwzględniono w niej także aspekt czasu oraz niepewności. Druga z opisanych ontologii została utworzona przez Autorkę artykułu. Powstała poprzez adaptację metod używanych przez historyków do weryfikacji informacji. Artykuł zawiera także wskazanie podobieństw oraz różnic w zaprezentowanych ontologiach.

**Słowa kluczowe:** sieć semantyczna, ontologie, ocena wiarygodności.

### 1. Wstęp

Ocena informacji pozyskiwanej z Internetu nie jest łatwym zadaniem. W ostatnich latach nastąpił gwałtowny rozwój portali opartych na Web 2.0, które są tworzone przez użytkowników. Przykładem takiego portalu jest Wikipedia – wolna internetowa encyklopedia. Każdy użytkownik sieci ma możliwość tworzenia i edycji haseł, których poprawność nie jest sprawdzana. Fałszywe wpisy mogą być dostępne dla innych użytkowników nawet przez wiele miesięcy. Sprawdzenie wiarygodności informacji może uchronić przedsiębiorstwo przed podjęciem błędnej i szkodliwej dla niego decyzji. Dlatego też zagadnienie weryfikacji informacji i ich źródeł w Internecie staje się ważnym zagadnieniem badawczym i praktycznym.

Artykuł przybliży problematykę oceny wiarygodności informacji pozyskanej z Internetu. Zawarto w nim także charakterystykę dwóch semantycznych podejść do jej oceny oraz ich porównanie. Pierwsze z podejść zostało zaproponowane przez Foxa i Huanga, natomiast drugie przez Autorkę artykułu.

### 2. Ocena wiarygodności informacji pozyskanej na potrzeby zarządzania

Wzorem dla wykorzystywanych obecnie sposobów badania wiarygodności informacji pozyskiwanych z Internetu są systemy kontroli informacji drukowanych. In-

ternetowa informacja uznawana jest za wiarygodną, jeśli [Batorowska, Kamińska-Czubala 2002, s. 92-93]:

- 1) została opublikowana w sieci przez autorytatywne źródło, np. organ rządowy lub samorządowy, znaną instytucję lub organizację,
- 2) stanowi elektroniczny odpowiednik, część lub zapowiedź drukowanego dokumentu – wówczas jest akceptowana od razu jako pewna,
- 3) jest oceniana przez ekspertów z danej dziedziny wiedzy.

Ocena jakości i wartości dotyczy każdej informacji wykorzystywanej w procesie zarządzania. W przypadku wykorzystania informacji i dalszego nią operowania istotne znaczenie ma ustalenie jej wiarygodności [Oleander-Skorek, Wydro 2007]. Wiarygodność jest zatem jednym z podstawowych kryteriów oceny jakości informacji, a w konsekwencji także wartości informacji.

Wiarygodność to stopień, w jakim ludzie mają zaufanie do danego wydarzenia, danej rzeczy [Kołaczek 2004]. Synonimami wiarygodności są prawość, godność i zaufanie. Te ostatnie można rozumieć jako: przekonanie, zamiar, oczekiwanie, uczucie, stosunek.

Wiarygodność informacji i jej źródła to problem, który w dużej mierze dotyczy podmiotów, dla których Internet jest prawdziwą bramą do świata realnego. Są to przede wszystkim sklepy i banki internetowe. Szczególnie ważna jest opinia o podmiocie, gdy do elementów wymiany należy m.in. pieniądź. Wtedy istotna staje się wiarygodność oferenta. Firmy, chcące umocnić swoją wiarygodność, wymyślają szereg działań, które najczęściej wiążą się z identyfikacją samych internautów [Lisewska 2004].

Wiarygodność informacji stała się zatem dla współczesnych menedżerów istotnym problemem, wynikającym z niekorzystnego zjawiska, jakim jest publikowanie błędnych informacji, a nawet ich fałszowanie. Powoduje to w konsekwencji naruszanie wizerunku wiarygodnej informacji publikowanej w Internecie, stanowiącej źródło wiedzy dla organizacji gospodarczych.

## 2.1. Trudności oceny wiarygodności informacji publikowanej w Internecie

Ze względu na specyficzny charakter informacji bardzo ważne staje się badanie jej wiarygodności, a co za tym idzie – również badanie wiarygodności źródła, z którego ona pochodzi. Wielość informacji w Internecie może jednak powodować trudności w jej ocenie oraz wpływać na małą użyteczność stron WWW jako źródła informacji dla zarządzania organizacją. Mogą to powodować następujące zdarzenia [Lisewska 2004]:

- nadmiar rezultatów wyszukiwania danej informacji,
- publikowanie informacji przez nieznanne podmioty – w Internecie źródłem informacji może być każda strona WWW i każdy mniejszy serwis,
- podszywanie się niektórych podmiotów pod inne poprzez tworzenie serwisów podobnych graficznie – można tu także wyróżnić działania związane z nieuczciwą konkurencją,

- niedostępność pewnych danych – dane te są dostępne odpłatnie lub wcale,
- różnego rodzaju błędy, jakie pojawiają się w nawet najlepiej prowadzonych serwisach WWW, od ortograficznych poprzez merytoryczne, np. pomyłki w wydarzeniach czy faktach.

Selekcja informacji według kryteriów jakościowych wymaga, by autor informacji był bardzo dobrze zorientowany w danej dziedzinie. Autor jest jednym z filtrów, który stosuje się przy czytaniu. Waga informacji zależy od jej źródła i oczywiście informacja jest tym cenniejsza, im wyżej ceni się w danej organizacji jej źródło. Wiadomo bowiem, że pewnym organizacjom, ze względu na ich potęgę ekonomiczną czy polityczną, przynależy szczególnie autorytet i szacunek, a w konsekwencji do ich publikacji lub stron internetowych przykładają się wielką wagę. Dotyczy to na przykład rządu, Unii Europejskiej, uniwersytetów [Nicholas 2001, s. 58].

## 2.2. Kryteria i zasady oceny wiarygodności informacji

Informację znaną w Internecie można uznać za wiarygodną, jeżeli [Szpunar 2007, s. 50]:

1) można ją potwierdzić w kilku niezależnych źródłach – jeśli o informacji pisze kilka serwisów, to można założyć, że jest ona prawdziwa,

2) podmiot publikujący jest autorytetem – podmiot udostępniający informacje ma swoje odbicie w rzeczywistości materialnej albo cieszy się społecznym zaufaniem, albo też szeroko znany jest autorytet autora,

3) źródło informacji jest stroną rządową (z rozszerzeniem gov), mimo że stosunkowo rzadko aktualizowana, można uznać za wiarygodną.

Użyteczny może być też zbiór zasad przyjętych co prawda w weryfikacji informacji medycznej, ale wydaje się, że może on być zastosowany też w przypadku innych typów informacji. Informacja medyczna, według Komisji Europejskiej, aby była uznana za wiarygodną, powinna spełniać następujące zasady [Bednarek-Michalska 2007]:

- 1) transparentności i rzetelności,
- 2) autorytetu,
- 3) prywatności i ochrony danych,
- 4) aktualizacji informacji,
- 5) odpowiedzialności,
- 6) dostępności.

Istotą pierwszej z wymienionych zasad jest określenie wszystkich danych adresowych dostawców treści, a także celów i zamiarów grupy, do której skierowana została dana informacja. Dzięki zasadzie transparentności i rzetelności możliwa jest pełna jawność autora [Bednarek-Michalska 2007].

Dzięki zasadzie autorytetu możliwe jest posiadanie jasnego statusu wykorzystywanych źródeł informacji i danych w bazach. Uzyskuje się wiarygodne dane o instytucjach opracowujących dane i daty ich dostarczenia [Bednarek-Michalska 2007].

Zasada trzecia dotyczy prywatności danych, polegającej na poddaniu informacji polityce zabezpieczania danych oraz procedurom ich przetwarzania [Bednarek-Michalska 2007].

Dzięki zasadzie aktualizacji informacji zapewniona jest bieżąca aktualizacja stron, o czym świadczą daty wprowadzenia zmian. Jeśli na danej stronie są linki, monitorowana jest również aktualność witryn, do których one odsyłają [Bednarek-Michalska 2007].

Kolejna zasada jasno określa odpowiedzialność za informacje. Podane są nazwiska ekspertów, którzy je przygotowali. Dokładnie przedstawieni są partnerzy współpracy oraz zapewniona jest ich wiarygodność. Jasno opisana została polityka edytowania stron i doboru materiału [Bednarek-Michalska 2007].

Zasada dostępności ma na celu zapewnienie użytkownikom łatwego dostępu do źródeł w postaci prostej nawigacji oraz poprawnego wyszukiwania [Bednarek-Michalska 2007].

Przedstawione zasady i kryteria nie rozwiązują w pełni problemu oceny wiarygodności informacji, a w szczególności oceny wiarygodności informacji wyszukiwanej przez agentów programowych. W związku z tym podjęto próby opracowania semantycznych mechanizmów oceny informacji znalezionej w Internecie.

### 3. Podejście zaproponowane przez M.S. Foxa i J. Huanga

Zagadnieniem, które zainteresowało Foxa i Huanga [2010], a które związane jest z wiarygodnością informacji pozyskanej z Internetu, jest pochodzenie wiedzy. Zwracają oni uwagę na sposób oceny prawdziwości informacji za pomocą modelowania i zachowania źródeł informacji, zależności między nimi i struktur zaufania. Fox i Huang wyróżnili cztery poziomy pochodzenia wiedzy:

- 1) statyczne,
- 2) dynamiczne,
- 3) zorientowane na niepewność,
- 4) oparte na osądach.

W celu określenia pochodzenia wiedzy należy odpowiedzieć sobie na następujące pytania:

- Czy można daną informację uznać za prawdziwą?
- Kto utworzył informację?
- Czy można uznać twórcę za godnego zaufania?
- Od czego to zaufanie zależy?
- Czy uznanie informacji za prawdziwą lub fałszywą może zależeć od oceny autora?

Fox i Huang zwracają także uwagę na duże znaczenie podpisu elektronicznego i cyfrowej certyfikacji w sieci zaufania. Należy jednak pamiętać, że zapewniają one tylko zbliżenie się do identyfikacji autora informacji, natomiast nie determinują jej wiarygodności. W kontekście pochodzenia wiedzy mogą one być używane tylko

do określenia twórcy informacji, określenia, czy informacja jest taka sama, jak ją oryginalnie zdefiniowano. Nie mogą natomiast wskazywać, czy informacja jest wiarygodna.

Opierając się na trzech pierwszych poziomach pochodzenia wiedzy, Fox i Huang opracowali model ontologii. W tym celu wykorzystali metodykę zaproponowaną przez Gruningena i Foxa. Ontologia ta tworzona była stopniowo: od poziomu statycznego, poprzez rozszerzenie o elementy należące do poziomu dynamicznego, do uzupełnienia o stopień niepewności. W dalszej części artykułu zostaną opisane poszczególne fazy ontologii. Poziom czwarty pochodzenia wiedzy został – tak samo jak w ontologii – pominięty.

### 3.1. Statyczny poziom pochodzenia wiedzy

Statyczne badanie pochodzenia koncentruje się na wiedzy, która jest pewna i nie zmienia się w czasie. Jest to najprostsza z występujących form pochodzenia. W celu dostarczenia formalnej i jednoznacznej specyfikacji statycznego pochodzenia wiedzy i do udostępnienia jej w Internecie Fox i Huang stworzyli „ontologię statycznego pochodzenia wiedzy”.

Każda informacja zbudowana jest ze zdania lub zdań. Zdanie może być proste (twierdzenie) i złożone (składające się z przesłanki i wniosku). Zdanie to najmniejszy kawałek informacji, któremu atrybuty powiązane z pochodzeniem mogą być przypisane. Może ono być prawdziwe lub fałszywe. Badając każde zdanie, należy zadać sobie następujące pytania:

- Czy zdanie jest prawdziwe?
- Kto jest autorem zdania?
- Czy badający powinien wierzyć osobie lub organizacji, która wytworzyła zdanie?
- Co jest podstawą dla badającego, by uznać zdanie za prawdziwe?
- Jak można sformalizować proces badania pochodzenia informacji?

Przykładowo zdanie: „Zwłoka dłuższa niż 1 minuta w odebraniu telefonu może powodować, że klient będzie nieusatysfakcjonowany” może być uznane za prawdziwe, jeżeli zostało utworzone przez emerytowanego menedżera ds. obsługi klientów, który uważany jest za autorytet w tym zakresie.

W przypadku zdań złożonych można wyróżnić dwa typy zależności:

- prawda przesłanki zależy od prawdy twierdzenia znalezionej w innym dokumencie WWW;
- prawda wniosku zależy od prawdy przesłanki i od rozumowania które doprowadziło do wyciągnięcia wniosków.

Powszechne jest zamieszczanie informacji w kilku dokumentach. Reprodukacja zdania w drugim dokumencie prowadzi do równoważności relacji pomiędzy nimi. Jednakże relacja ta jest także asymetryczna – jedno zdanie jest kopią drugiego. Kopia zdania jest klasyfikowana jako zdanie równoważne. Ponadto zdanie może być wyprowadzane za pomocą logiki dedukcyjnej.

Każda ontologia zawiera także terminy i aksjomaty niezbędne do odpowiedzi na pytania kompetencyjne. W przypadku ontologii dla statycznego poziomu pochodzenia wiedzy pytaniami tymi są:

- Czy dane zdanie jest prawdziwe, fałszywe lub niemożliwe do określenia?
- Kto jest autorem zdania?
- Jaki jest status weryfikacji podpisu elektronicznego?
- Jakiego zakresu wiedzy dotyczy zdanie?
- Czy w tym zakresie wiedzy twórca informacji może być uznany za osobę godną zaufania?
- Czy prawdziwość badanego zdania zależy od innego zdania? Jeśli tak, to od jakiego?

W ontologii statycznego pochodzenia wiedzy Fox i Huang wyróżnili pięć klas:

- twierdzenia,
- dokumenty Web,
- źródła informacji,
- relacje zaufania,
- status podpisu elektronicznego.

W ontologii tej określono także relację zaufania, która definiuje, czy badający pochodzenie ufa, jakim twierdzeniom ufa i w jakich kwestiach. Relacja zaufania w pochodzeniu wiedzy definiowana jest jako trójka  $(a, c, f)$ , gdzie badający pochodzenie (odbiorca informacji)  $a$  ufa twórcy informacji  $c$  w kwestii lub w zakresie specjalistycznego pola wiedzy  $f$ . „Ufa” oznacza, że uznaje za prawdziwe jakiekolwiek twierdzenie utworzone przez  $c$  w obszarze  $f$ .

Zbiór aksjomatów jest zdefiniowany w celu określenia stanu prawdziwości zdania. Wiarygodność zależy od tego, czy zdanie jest „godne zaufania” lub w przypadku zdania złożonego – od wiarygodności zdania, którego jest kopią lub z którego zostało wyprowadzone. W dodatku zdanie jest godne zaufania, jeśli twórca lub wydawca twierdzenia jest godny zaufania w jednym z obszarów, którego zdanie dotyczy, oraz weryfikacja podpisu elektronicznego jest poprawna.

Twórcy informacji potrzebują „zaopatrzyć” dokumenty internetowe metadany- mi ontologii statycznego pochodzenia wiedzy. Natomiast użytkownicy muszą zdefiniować relacje zaufania, a moduł wnioskujący musi przeprowadzić wnioskowanie na opisanych dokumentach WWW.

### 3.2. Dynamiczne pochodzenie wiedzy

Jednakże powszechnie wiadomo, że w prawdziwym świecie informacja może zmieniać się w czasie. Przykładowo w łańcuchu dostaw ceny, inwentarz, powierzchnia magazynowa zmieniają się w czasie. Informacje uznane za wiarygodne w danym okresie czasu po jego upływie mogą być uznane za niewiarygodne. Tak samo relacja zaufania może się zmieniać w czasie i autor informacji może być godny zaufania tylko w danym okresie czasu. Okres, w którym jest on wiarygodny, nazywa się „efektywnym okresem”.

Rozszerzając podaną w poprzednim punkcie definicję relacji zaufania ( $a$ ,  $c$ ,  $f$ ), można stwierdzić, że agent pochodzenia  $a$  ufa twórcy twierdzenia  $c$  w obszarze wiedzy  $f$ , może mieć także efektywny okres. Relacja zaufania nazywana jest efektywną w podanym punkcie czasu, jeśli punkt ten jest w okresie efektywnym relacji zaufania.

Zatem zdanie jest wiarygodne w podanym punkcie czasu, jeśli jego twórca jest godny zaufania w jednym z obszarów twierdzenia w punkcie czasu oraz weryfikacja podpisu elektronicznego jest poprawna. Natomiast twórca informacji jest wiarygodny w określonym obszarze wiedzy w podanym punkcie czasu, jeśli punkt czasu należy do okresu efektywnego w relacji zaufania.

### 3.3. Pochodzenie wiedzy zorientowane na niepewność

Ocenę wiarygodności informacji często komplikuje fakt, że jej autorowi można raczej zaufać niż ufać mu całkowicie lub całkowicie nie ufać. Ponadto utworzona przez autora informacja może być raczej prawdziwa w określonym stopniu niż absolutnie prawdziwa lub absolutnie fałszywa.

Rozszerzając ontologię, Fox i Huang wprowadzili pojęcia **stopnia zaufania** i **stopnia wiary**. Pierwszy z nich jest subiektywnym prawdopodobieństwem wiarygodności informacji. Może przyjąć wartości od 0 do 1. Natomiast stopień wiary jest subiektywnym prawdopodobieństwem reprezentującym stopień zaufania niepewnych relacji zaufania.

## 4. Propozycja ontologii opartej na adaptacji metod historycznych do oceny wiarygodności informacji

Podejście zaproponowane przez Autorkę [Kajfosz 2009a, 2009b] artykułu oparte jest na metodach wykorzystywanych przez historyków. Na podstawie ich analizy Autorka wskazała, że badając wiarygodność informacji zamieszczonych w Internecie, należy wziąć pod uwagę czynniki podobne jak przy badaniu źródła historycznego:

- określenia czasowe: data wydarzenia, data utworzenia, data modyfikacji,
- miejsce utworzenia,
- autora,
- źródło publikacji informacji,
- powtarzalność informacji.

Jednakże wydaje się także celowe rozbudowanie metody historycznej o badanie następujących czynników:

- motywy publikacji informacji,
- źródła zewnętrzne,
- popularność źródła.

Ocena wiarygodności informacji pozyskiwanej z Internetu jest tak samo jak krytyka historyczna procedurą. Wzorując się na procedurze krytyki historycznej, proponuje się przyjąć następujące etapy:



- określenie czasu powstania informacji i jej aktualności,
- określenie miejsca powstania informacji,
- ustalenie autora,
- ustalenie celu publikacji badanej informacji,
- określenie, czy badana informacja pojawia się w innych źródłach,
- badanie źródła, z którego pochodzi informacja wraz z określeniem jego typu i popularności,
- określenie wykorzystywanych źródeł zewnętrznych.

Pierwszych pięć etapów dotyczy oceny informacji, natomiast ostatnie dwa poświęcone są weryfikacji jej źródła.

Określenie czasu powstania informacji polega na ustaleniu daty wydarzenia, utworzenia i ostatniej modyfikacji. Data wydarzenia wskazuje na czas, w jakim opisywane wydarzenie miało miejsce. Data utworzenia informuje, kiedy autor utworzył informację. Jeśli data utworzenia jest w znacznie późniejsza niż data wydarzenia, można wysnuć hipotezę o mniejszej wiarygodności źródła. Wiadomo że po upływie czasu ludzki umysł ma właściwość zniekształcania informacji, pomijania istotnych szczegółów, które z biegiem czasu przestaje uważać za istotne. Natomiast data ostatniej modyfikacji wskazuje na aktualność źródła. Przeszukując Internet, użytkownicy odnajdują często dane i informacje sprzed kilku lat, trafiają na strony WWW dawno nieaktualizowane, które powinny być z sieci usunięte. W przypadku takich informacji i ich źródeł można wysnuć przypuszczenie, że są one już nieaktualne. Przy dzisiejszym tempie rozwoju społecznego nawet informacje kilkumiesięczne mogą już być nieistotne i niewarte trudu dalszych badań. Należy jednak pamiętać, że pojęcie aktualności jest pojęciem względnym. W przypadku niektórych typów informacji (np. finansowych) aktualność jest bardzo ważna. Natomiast w ocenie wiarygodności innych informacji (np. historycznych) odgrywa małe znaczenie.

Drugim etapem badania wiarygodności informacji pozyskiwanej z Internetu jest określenie miejsca jej powstania. W przypadku informacji internetowych większość wiadomości dotyczących miejsca utworzenia można odczytać z adresu URL. Analizując adres URL, można określić kraj, region czy nawet miasto, gdzie informacja została utworzona, oraz charakter organizacji, która informację opublikowała. Organizacje można uszeregować według stopnia wiarygodności w następujący sposób:

- rządowe (gov),
- międzynarodowe (int),
- edukacyjne (edu),
- militarne (mil),
- niekomercyjne (org),
- związane z siecią (net),
- komercyjne (com).

Najbardziej wiarygodne są informacje pochodzące ze stron organizacji rządowych. Należy jednak pamiętać, że serwisy znajdujące się na stronach gov są rzadko aktualizowane przez co informacja może być przestarzała.



Kolejnym etapem opisywanej procedury jest ustalenie autora informacji, którym może być zarówno osoba fizyczna, jak i przedsiębiorstwo. Autora informacji i jej źródło można także odczytać z adresu URL lub bezpośrednio ze strony internetowej. Jeśli autorem jest osoba fizyczna, należy określić jej imię, nazwisko oraz nick używany w Sieci, a w przypadku przedsiębiorstwa – jego nazwę. Należy pamiętać, że w przypadku źródeł internetowych podszywanie się pod inną osobę jest bardzo proste. Nie jest sprawdzana autentyczność osób podpisujących się pod daną wiadomością. Oprócz tego istnieje wiele możliwości przechwycenia loginu i hasła do serwisów wymagających uwierzytelniania. Osoba posiadająca przechwycone dane może zamieścić na takiej stronie fałszywe informacje pod nazwiskiem specjalisty z danej dziedziny, powołując się na jego autorytet.

Badając autora źródła, należy sprawdzić jego uprawnienia i kwalifikacje (m.in. posiadany stopień naukowy, specjalizację badawczą, praktykę zawodową, zdobyte odznaczenia i nagrody). Za bardziej wiarygodne uważa się osoby posiadające wysoki tytuł naukowy, będące specjalistami w danej dziedzinie, mające wybitne osiągnięcia lub wyróżnione prestiżowymi nagrodami. Analizując autora informacji, należy określić jego obiektywizm. Autor emocjonalnie związany z treścią informacji jest mniej wiarygodny. Tak samo jest w przypadku autora, który nie brał udziału w opisywanych wydarzeniach. Na tym etapie badania wiarygodności należy także określić odnośniki do strony WWW autora oraz do strony WWW firmy/firm, z którą autor jest związany. Wiarygodność autora zwiększa podanie danych kontaktowych. Daje to możliwość jego weryfikacji oraz bezpośredniego kontaktu z autorem (np. za pomocą poczty e-mail).

Kolejnym ważnym etapem procesu oceny wiarygodności informacji pozyskiwanych z Internetu jest ustalenie celu zamieszczenia informacji na stronie. Informacje, których cel jest zgodny z celem strony, uważane są za bardziej wiarygodne.

Należy również sprawdzić, czy badana informacja pojawiła się także w innych źródłach – zarówno tych internetowych, jak i papierowych. Następnie należy sprawdzić, czy źródła te są z sobą powiązane. Kolejnym krokiem jest ustalenie źródeł, z jakich autorzy korzystali. Informacje, których autorzy korzystali z tego samego źródła, są mniej wiarygodne.

Badając źródło informacji, należy zwrócić uwagę na obecność podstrony „O Nas”/„O Stronie”. Powinien znajdować się tam jasno określony cel powstania strony. Na wiarygodność źródła informacji wpływa także szata graficzna strony. Powinna być zaprojektowana w sposób profesjonalny, przejrzysty i czytelny. Strona także powinna być łatwa w użyciu i przyjazna dla użytkownika. Należy zbadać również, czy na stronie znajdują się odsyłacze do innych stron WWW, następnie określić ich tematykę i autorytet.

Podczas weryfikacji źródła informacji należy określić typ strony. Źródło informacji może okazać się:

- blogiem,
- forum dyskusyjnym,

- stroną firmową,
- stroną konkurencji,
- stroną sponsorowaną,
- stroną niezależną.

Należy pamiętać, że informacje zamieszczane na blogach z reguły mają charakter osobistych przemyśleń autora i są subiektywne. Także do informacji pochodzących z forów dyskusyjnych należy podchodzić z dystansem. Informacje zamieszczone na stronie internetowej firmy albo na stronie sponsorowanej mogą mieć charakter reklamowy, nie do końca zgodny z opisywaną rzeczywistością. Informacje pochodzące ze strony konkurencji mogą posiadać negatywny oraz subiektywny charakter. Najbardziej wiarygodne są więc informacje umieszczane na niezależnych serwisach internetowych. Można przypuszczać, że taka informacja jest obiektywna.

Weryfikując źródło, należy także zbadać jego popularność. Należy sprawdzić, czy istnieją odnośniki z innych stron oraz z katalogów znanych portali. Warto zapoznać się również z różnego rodzaju statystykami (np. licznik odwiedzin), które wskazują na popularność badanego źródła w środowisku internetowym. Pokreślić należy fakt, iż popularność źródła nie może być jedynym czynnikiem wskazującym na wiarygodność informacji z niego pochodzącej. Strony, które są popularne w pewnych środowiskach (np. Wikipedia wśród uczniów), mogą dostarczać błędnych informacji.

Na wiarygodność informacji wpływa także wskazanie źródeł zewnętrznych wykorzystywanych przy tworzeniu informacji (wraz z określeniem ich autora, tytułu i miejsca w sieci, gdzie się znajdują). Ważne jest określenie, czy badane źródło jest źródłem pierwotnym, ponieważ źródła te darzy się większym zaufaniem. Jeśli badane źródło nie jest pierwotne, należy sprawdzić, czy zawiera ono odwołanie do takiego źródła.

Opierając się na powyższej procedurze, Autorka utworzyła ontologię, której celem jest dostarczenie agentom poszukującym informacji w Internecie możliwości oceny ich wiarygodności. Ma to zapobiec dodawaniu do baz wiedzy informacji niewiarygodnych lub o wątpliwej wiarygodności. Powstała ontologia powinna być na tyle uniwersalna, żeby można za jej pomocą opisać każdą informację pojawiającą się w sieci.

Ontologia wspomagająca badanie wiarygodności informacji pozyskiwanej z Internetu jest ontologią zadaniową, ponieważ dotyczy konkretnego działania – oceny wiarygodności informacji. W przypadku problemu oceny wiarygodności informacji pozyskiwanej z Internetu istnieje konieczność wprowadzenia mechanizmu wnioskowania, który dostarczy agentom poszukującym informacji w Internecie jej ocenę. Przy tworzeniu ontologii nie wykorzystano innej istniejącej ontologii.

Podstawowymi klasami w modelowanej ontologii są:

- *Informacja*,
- *WiarygodnoscTresci*,
- *WiarygodnoscZrodla*,
- *WiarygodnoscInformacji*.

Klasa *Informacja* charakteryzuje badaną informację. Jest ona powiązana z klasą *DataWydarzenia*, która zawiera dane o dacie opisywanego wydarzenia. Każda informacja może mieć przypisaną datę wydarzenia lub nie. Każda data wydarzenia może odpowiadać jednej lub większej liczbie informacji. Z klasą *Informacja* wiąże się także klasa *DataUtworzenia*, która wskazuje na datę utworzenia badanej informacji. Każda informacja ma dokładnie jedną datę utworzenia, natomiast każda data utworzenia może być przypisana kilku informacjom. Kolejną klasą, z którą łączy się *Informacja*, jest *DataAktualizacji*. Klasa ta wskazuje na aktualność badanej informacji. Wszystkie informacje powinny posiadać jedną (ostatnią) datę aktualizacji. Data aktualizacji może odpowiadać jednej lub większej liczbie informacji. Klasa *Powtarzalność* jest także powiązana z *Informacją*. Klasa ta wskazuje na to, czy badana informacja pojawia się także w innych źródłach. Może ona nie być publikowana w żadnym innym źródle lub ukazywać się w jednym, lub wielu źródłach. Każde źródło może publikować po jednej lub więcej informacji. W charakterystyce informacji ważny jest także jej autor, którego opisuje klasa *Autor*. Informacja ma jednego autora lub kilku autorów. Jeden autor może utworzyć jedną lub wiele informacji. Oprócz autora informacja powinna mieć określone także źródło, w którym została zamieszczona sama lub z innymi informacjami. Sytuację tę obrazuje powiązanie klasy *Informacja* z klasą *Zrodło*. Ostatnią klasą powiązaną bezpośrednio z klasą *Informacja* jest klasa *CellInformacji*, która opisuje cel, jaki przyświecał twórcom lub twórcom – jest to relacja jeden do jeden.

Klasa *Powtarzalność* jest także agregatem dwóch klas *MiejscePowtorzenia* i *Bibliografia*. Pierwsza z nich charakteryzuje źródło, w którym informacja została powtórzona. Druga klasa zawiera informacje o bibliografii źródła, w którym informacja została powtórzona.

Klasa *Autor* jest powiązana agregacją całkowitą z klasą *Odnosnik*, która zawiera informacje o odnośniku związanym z autorem informacji. Klasa zawierająca informacje o autorze jest także uogólnieniem dwóch klas *Przedsiębiorstwo* i *Osoba*. Pierwsza z nich charakteryzuje przedsiębiorstwo, które jest autorem informacji. Druga opisuje osobę, która jest autorem informacji. Klasa ta jest także agregatem dla klasy *Ekspert*, która charakteryzuje uprawnienia i kwalifikacje autora – osoby.

Klasa *Zrodło* jest uogólnieniem dla klas *ZrodłoPierwotne* i *ZrodłoWtorne*. Pierwsza z nich opisuje źródło, które jest pierwotne. Druga charakteryzuje źródło informacji, które jest wtórne.

Każde źródło internetowe posiada dokładnie jeden adres URL, który odpowiada dokładnie jednej stronie WWW. Klasa ta jest agregatem następujących klas:

- *Kraj* – charakteryzuje kraj, w którym informacja została wytworzona,
- *Region* – opisuje region, w którym informacja powstała,
- *Miasto* – zawiera dane o mieście, które było miejscem powstania informacji,
- *TypOrganizacji* – charakteryzuje typ organizacji, na której stronie informacja została opublikowana.

Ze *Zrodlem* wiąże się także klasa *TypZrodla*. Każde źródło jest jednego, określonego typu. *TypZrodla* jest uogólnieniem dla następujących klas – typów stron WWW:

- *Blog* – charakteryzuje źródło, które jest blogiem,
- *StronaSponsorowana* – opisuje stronę sponsorowaną, która jest źródłem informacji,
- *StronaNiezalezna* – zawiera dane o źródle – stronie niezależnej,
- *StronaFirmy* – zawiera charakterystykę miejsca na stronie firmy, na której informacja została opublikowana,
- *StronaKonkurencji* – zawiera opis typu źródła, jakim jest strona konkurencji,
- *Forum* – charakteryzuje forum, które jest źródłem informacji.

Klasa *ZrodloZewnetrzne* charakteryzuje stronę, do której odwołuje się badane źródło informacji. Klasa *SzataGraficzna* opisuje szatę graficzną analizowanego źródła. Klasa *CelZrodla* zawiera opis celu, dla którego źródło zostało utworzone. Klasa *InfoOZrodle* przedstawia dostępne informacje o źródle, z którego informacja pochodzi.

Klasa *Popularnosc* określa popularność analizowanego źródła. Jest ona agregatem dla następujących klas: *LicznikOdwiedzin*, *InnaStatystyka* i *Odsylacz*. Pierwsza z nich zawiera dane o liczbie odwiedzin na badanej stronie. Druga określa inną dostępną dla strony statystykę. Trzecia klasa zawiera informacje o odnośnikach z innej strony (klasa *ZInnejStr*) do innej strony (klasa *DoInnychStron*), z katalogów stron (klasa *KatalogStron*).

Klasa *WiarygodnoscTresci* zawiera częściową ocenę wiarygodności i dotyczy treści badanej informacji. Klasa ta łączy się z relacją jeden do jeden z opisanymi powyżej klasami: *DataWydarzenia*, *DataUtworzenia* i *DataAktualizacji*.

*WiarygodnoscZrodla* jest klasą zawierającą ocenę wiarygodności źródła, z którego informacja pochodzi. Na ocenę tę wpływają następujące klasy: *Zrodlo*, *URL*, *TypZrodla*, *ZrodloZewnetrzne*, *SzataGraficzna*, *CelZrodla*, *InfoOZrodle*, *Popularnosc*. Dla każdego źródła informacji można określić jego wiarygodność. Jeden obiekt klasy *WiarygodnoscZrodla* może odpowiadać więcej niż jednemu źródłu. Na proces badania wiarygodności źródła badanej informacji wpływa kilka czynników. Pierwszym z nich jest liczba źródeł zewnętrznych. Każde źródło może nie korzystać z żadnego lub korzystać z nieograniczonej ich liczby. Z każdą internetową stroną związane jest pojęcie szaty graficznej, które wpływa na ocenę jej wiarygodności. Jest to relacja jeden do jeden. Na wiarygodność źródła wpływa także cel jego powstania. Strona może mieć jasno sprecyzowany cel lub nie, może on być prosty lub złożony – charakteryzuje to liczebność zero lub wielu. Na każdej stronie powinny być umieszczone „Informacje o stronie”/”O nas” (zero lub jedna). Ostatnim wyróżnionym na diagramie ERD czynnikiem wpływającym na wiarygodność źródła jest jego popularność.



Najważniejszą klasą jest *WiarygodnoscInformacji*. Zawiera ona określenie wiarygodności badanej informacji. Agreguje ona klasy *WiarygodnoscTresci* oraz *WiarygodnoscZrodla*. Wpływają na nią następujące klasy: *CellInformacji*, *Autor*, *Powtarzalnosc*. Jest ona powiązana z klasą *Informacja* relacją wiele do jeden: dla każdej informacji można określić jedną wartość wiarygodności, dana wartość może odpowiadać kilku informacjom. Diagram ERD modelowanej ontologii przedstawia rysunek 1.

## 5. Porównanie opisanych podejść

Opisane w artykule ontologie posiadają pewne elementy wspólne, na które warto zwrócić uwagę. Podczas oceny informacji uwzględniają one jej autora. Zalecają także określenie zaufania, jakie badający ma w stosunku do autora. Na zaufanie wpływa w szczególności autorytet autora informacji w danej dziedzinie. W ontologiach tych zawarty jest również aspekt wielokrotnego publikowania tej samej informacji w kilku źródłach. Zarówno Fox i Huang, jak i Autorka artykułu wskazują, że informacje te są powiązane i należy je oceniać wspólnie. Obydwie ontologie uwzględniają także aspekt czasu. Fox i Huang wprowadzili do ontologii pojęcie „okresu efektywnego”, natomiast Autorka artykułu utworzyła klasę *DataAktualizacji*, której celem jest określenie, czy badana informacja jest aktualna.

Opisane podejścia różnią się między sobą. Pierwszą łatwo zauważalną różnicą jest sposób tworzenia ontologii. Fox i Huang rozpoczęli budowę ontologii od wyróżnieniach poziomów złożoności pochodzenia wiedzy. Natomiast Autorka artykułu najpierw opracowała, w oparciu o metody wykorzystywane przez historyków, procedurę oceny wiarygodności informacji, a następnie na jej podstawie opracowała ontologię. Opisane ontologie różnią się także wykorzystaną metodologią do ich budowy: Fox i Huang wykorzystali metodologię TOVE, natomiast Autorka artykułu – metodologię Noya i McGuinnessa. Kolejną bardziej istotną różnicą jest uwzględnienie przez Foxa i Huanga stopnia niepewności podczas badania wiarygodności informacji oraz elementów teorii prawdopodobieństwa. Autorka artykułu nie uwzględniła w ontologii prawdopodobieństwa, wykorzystowała natomiast skalę dyferencjału semantycznego w zaproponowanej przez nią procedurze oceny wiarygodności (por. [Kajfosz 2009a]). Ontologie różnią się także ilością elementów wpływających na wiarygodność informacji. W przypadku podejścia zaproponowanego przez Foxa i Huanga na prawdziwość informacji wpływa: autor, czas oraz powielenie informacji. Autorka artykułu uwzględniła w swojej ontologii wiele czynników mogących mieć wpływ na wiarygodność ontologii informacji pozyskanej z Internetu. Fox i Huang sporą wagę przykładają także do podpisu elektronicznego i statusu jego weryfikacji. Aspekt ten został pominięty przez Autorkę artykułu ze względu na stosunkowo małą popularność podpisu elektronicznego w Polsce.



## 6. Podsumowanie

Opisane w artykule ontologie służące do oceny wiarygodności informacji pozyskanej z Internetu, mimo że dotyczą tego samego zagadnienia, prezentują odmienne podejście do problemu. Fox i Huang stworzyli ontologię sformalizowaną, opartą na teorii prawdopodobieństwa. Natomiast Autorka artykułu stworzyła ontologię zawierającą więcej czynników wpływających na wiarygodność informacji. Wydaje się, że żadne z opisanych podejść nie jest jeszcze dostatecznie dopracowane.

Autorka artykułu w dalszych pracach badawczych rozwijać będzie stworzoną przez siebie ontologię. Warte rozważenia jest wprowadzenie do niej aspektu związanego z niepewnością – tak jak zrobili to Fox i Huang. Należy zastanowić się, czy nie lepiej użyć do tego celu logiki rozmytej. Określanie prawdopodobieństwa jest elementem subiektywnym.

## Literatura

- Batorowska H., Kamińska-Czubala B., *Informatyczna i dydaktyczna działalność szkolnego centrum informacji*, [w:] H. Batorowska, B. Kamińska-Czubala, *Szkolne centrum informacji*, Kraków 2002.
- Bednarek-Michalska B., *Ocena jakości informacji elektronicznej. Pułapki sieci*, 2007, <http://www.ebib.info/2007/86/a.php?bednarek> (20.07.2010).
- Fox M.S, Huang J., *Knowledge provenance in enterprise information*, International Journal of Production Research 2010, vol. 43, issue 20.
- Kajfosz K., *Koncepcja ontologii na potrzeby oceny wiarygodności informacji pozyskiwanej z Internetu*, [w:] J. Gołuchowski, A. Frączkiewicz-Wronka (red.), *Technologie wiedzy w zarządzaniu publicznym*, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice 2009b.
- Kajfosz K., *Ocena wiarygodności informacji pozyskiwanej z Internetu z wykorzystaniem metod semantycznych*, praca magisterska, Katowice 2009a.
- Kołaczek G., *Dynamika relacji zaufania i jej modelowanie w systemach teleinformatycznych*, [w:] A. Grzywka (red.), *Internet w społeczeństwie informacyjnym: podstawowe problemy Internetu*, Warszawa 2004.
- Lisewska J., *Internet – wiarygodne źródło informacji czy przypadkowych wiadomości*, [w:] Szewczyk A. (red.), *Informacja – dobra lub zła nowina*, Szczecin 2004.
- Nicholas D., *Ocena potrzeb informacyjnych w dobie Internetu*, Warszawa 2001.
- Oleander-Skorek M., Wydro K.B., *Wartość informacji*, „Telekomunikacja i Techniki Komunikacyjne” 2007, 1-2.
- Szpunar M., *Internet źródło informacji versus dezinformacji. O jakości i rzetelności informacji pozyskiwanych w Internecie*, „E-mentor” 2007, nr 2, s. 46-51.

## OUTLINE OF ONTOLOGIES FOR EVALUATION OF INTERNET-BASED INFORMATION TRUSTWORTHINESS

**Summary:** Information is very important in every activity of a company. For this reason using Internet to search for the information has quickly increased. This paper discusses a very important issue of the Internet-based information – its trustworthiness. It describes difficul-



ties, which are encountered during information trustworthiness verification, and evaluation methods, which have been developed. The paper presents also two ontologies to verify Internet-based information trustworthiness. The first one, created by Fox and Huang, focuses on knowledge provenance. In this ontology there are two things in which pressure was put on: an author and a reproduction of information in other sources. It also considers two aspects: time and uncertainty. The second of described ontologies was created by the paper's author. It was built by adapting historians methods to verify the information. The paper also contains similarities and differences in presented ontologies.