

DOI: 10.18559/SOEP.2017.12.6.

Jacek Małyшко, Dawid Wękowski

Uniwersytet Ekonomiczny w Poznaniu, Wydział Informatyki i Gospodarki Elektronicznej

Autor do korespondencji: Jacek Małyшко, jacek.malyszko@ue.poznan.pl

PROFILOWANIE UŻYTKOWNIKÓW Z WYKORZYSTANIEM TOŻSAMOŚCI ŹRÓDEŁ INFORMACYJNYCH

Streszczenie: Wraz ze wzmożoną aktywnością ludzi w sieci coraz więcej informacji na ich temat manifestuje się w tym, jakie przeglądają strony, jakie dane zamieszczają o sobie lub z kim się kontaktują. Takie informacje mogą posłużyć do budowy wirtualnych tożsamości użytkowników sieci. W niniejszej pracy przedstawiono wyniki badań nad problemem budowania początkowej tożsamości użytkownika w obliczu silnie ograniczonej kolekcji informacji o nim, dostępnej w pierwszych etapach generowania jego profilu. W pracy zaproponowano takie rozwiązanie tego problemu, w którym tożsamości początkowe konstruowane są na podstawie tożsamości źródeł internetowych, zadeklarowane przez użytkownika jako zawierające interesujące dla niego treści. W przeprowadzonym eksperymencie wykazano, że tożsamości bytów manifestujących się w internecie mogą posłużyć do budowania stereotypów tożsamości dla nowych użytkowników.

Słowa kluczowe: profilowanie użytkownika, systemy adaptacyjne, zarządzanie tożsamością.

Klasyfikacja JEL: C88, D83.

USER PROFILING BASED ON THE IDENTITIES OF INFORMATION SOURCES

Abstract: Along with the increased activity of people on the web, more and more information about them can be gathered based on the analysis of which web pages they visit, what data about themselves they put on the web and who they make contact with. Such information can be used to build so-called virtual identities of web users. This article presents the results of research on the problem of building an initial user identity in a situation, when the information about the user is limited very much, i.e. in the early stages of interaction with the system. A solution to this problem is proposed, in which the initial identities are constructed and based on the identity of information sources, which have been declared by the users as relevant to their information needs. The conducted experiment revealed that such identities of information sources can be used to build initial user identities.

Keywords: user modelling, adaptive systems, identity management.

Wstęp

Obecnie wiele czynności znanych dotychczas tylko ze świata rzeczywistego przenosi się do świata wirtualnego. Wraz ze wzmożoną aktywnością ludzi w sieci coraz więcej informacji na ich temat manifestuje się w tym, jakie przeglądają strony, jakie dane o sobie zamieszczają lub z kim się kontaktują. Takie informacje mogą posłużyć do budowy wirtualnych tożsamości użytkowników sieci, rozumianych jako ich profile, które mogą być wykorzystane w różnorodnych scenariuszach personalizacji usług elektronicznych.

W niniejszej pracy przedstawiono wyniki badań nad problemem budowania początkowej tożsamości użytkownika w obliczu silnie ograniczonej kolekcji informacji o nim, dostępnej w pierwszych etapach generowania profilu. Drugim obszarem problemowym podejmowanym w pracy jest możliwość przeniesienia pojęcia tożsamości na byty inne niż człowiek. Zaprezentowano podejście pozwalające na budowanie wirtualnych tożsamości dla źródeł informacyjnych dostępnych w internecie.

W części 1 przedstawiono analizę literatury w zakresie budowania tożsamości użytkowników w internecie. Część 2 opisuje system Ego, w ramach którego przeprowadzany był eksperyment. Opis podejścia związanego z tworzeniem tożsamości początkowych oraz tożsamości bytów innych niż człowiek znajduje się w części 3. Część 4 zawiera opis i ewaluację eksperymentu, a całość pracy kończy podsumowanie.

1. Analiza literatury

Pojęcie tożsamości w systemach informacyjnych odnosi się do formalnej reprezentacji wiedzy lub kolekcji twierdzeń formułowanych przez jeden podmiot, manifestujący się w świecie cyfrowym, o samym sobie lub o innej jednostce [Cameron 2005]. Podmiotem, o którym mowa, może być dowolny byt, jaki można opisać za pomocą atrybutów [Kubicek 2010], a więc osoba (np. użytkownik, pracownik, klient), byt abstrakcyjny (np. osoba prawna, grupa, organizacja), obiekt świata rzeczywistego (np. urządzenia, pojazdy) oraz obiekt wirtualny (np. serwisy internetowe, awatary) [Miorandi i in. 2012; Rundle i in. 2010]. W literaturze można znaleźć przeglądy różnych podejść do opisu i definicji tożsamości w kontekście systemów informacyjnych [Nabeth 2009; Clarke 2008]. Najczęściej jednak pojęcie tożsamości związane jest z budowaniem charakterystyki osoby za pomocą zestawu atrybutów, którym nadaje się określone wartości [Rannenberg i in. 2009]. Wśród wielu rodzajów takich atrybutów, można wyróżnić [Rannenberg, Royer i Deuker 2009; Nabeth 2009; Kubicek 2010]:

- atrybuty świata rzeczywistego, takie jak nazwisko, adres, zainteresowania, umiejętności, role społeczne, działania i miejsca w świecie rzeczywistym,
- atrybuty cyfrowe, takie jak połączenia w internetowych sieciach społecznych, historia przeglądanych stron internetowych, muzyka, filmy,
- atrybuty używane do uwierzytelniania i autoryzacji, takie jak loginy i hasła, prawa dostępu, dane biometryczne.

Wirtualna tożsamość nie może objąć wszystkich cech podmiotu, do którego się odnosi, dlatego tożsamości często uznaje się jedynie za częściową reprezentację bytu. Dodatkowo wirtualne tożsamości należy rozważać także w kontekście czasu i przestrzeni, w których są budowane i wykorzystywane. Tożsamość może być rozumiana jako stabilna, ewoluująca w czasie reprezentacja podmiotu lub jako opis stworzony ad hoc, na potrzeby pojedynczego użycia. Można również przyjąć, że tożsamość jest spójną reprezentacją danej osoby (użytkownika) albo jest silnie zależna kontekstu i w różnych sytuacjach manifestuje się w odmienny sposób, dostosowując się do wymogów i warunków [Rannenberg, Royer i Deuker 2009]. Jak widać, wirtualna tożsamość może być różnie definiowana w zależności od podmiotu, dla którego jest tworzona, sposobu opisu podmiotu oraz kontekstu wykorzystania.

Wraz z rozwojem koncepcji wirtualnych tożsamości użytkowników w internecie powstało wiele inicjatyw, projektów oraz rozwiązań z zakresu zarządzania tożsamością. Wśród protokołów umożliwiających budowanie i wykorzystywanie tożsamości w sieci można wyróżnić grupę protokołów silnie

związanych z uwierzytelnianiem użytkowników na portalach internetowych. Są to m. in. OpenID [OpenID 2007], Persona/BrowserID [Gilbert i Upatising 2013] czy WebID [Story i Corlosquet 2011]. Oprócz nich funkcjonują również protokoły wykorzystywane w procesie autoryzacji użytkowników, jak OAuth [Hammer-Lahav 2011], a także służące do przekazywania dowolnych twierdzeń odnoszących się do bytów cyfrowych, na przykład SAML [Ragouzis i in. 2006].

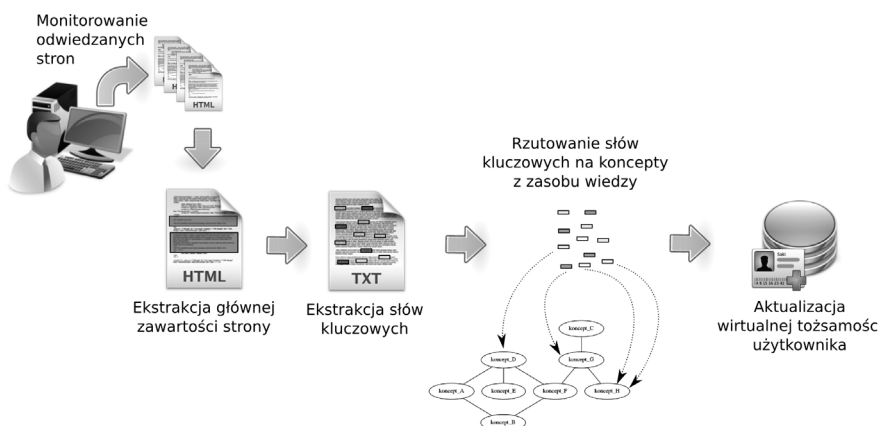
Wiele rozwiązań realizujących zadania z zakresu zarządzania tożsamością użytkownika związanych jest z różnego rodzaju projektami badawczymi. Znaczna część z nich skupia się na kwestii zachowania prywatności podczas udostępniania informacji w sieci oraz na wspomaganium użytkowników w zarządzaniu informacjami przekazywanymi określonym podmiotom w internecie. Rozwiązania te zorganizowane są na różnym stopniu ogólności – od zarządzania pojedynczymi danymi aż po złożone rozwiązania (*framework*). Projekt Personal [Personal 2013] pozwala na porządkowanie zestawów danych wykorzystywanych w sieci oraz automatyczne wypełnianie formularzy internetowych na podstawie tych danych. Projekty takie jak Higgins [Higgins 2013] wykorzystują tzw. karty informacyjne, czyli przygotowane zestawy informacji o użytkowniku, pokazujące pewien wycinek tożsamości, przeznaczony do zaprezentowania na określonych portalach. Z kolei Digital.Me [Scerri i in. 2011] to projekt umożliwiający budowę zdecentralizowanej sieci społecznej z wbudowanymi mechanizmami do sterowania udostępnianiem informacji. Projekt PrimeLife [Camenisch, Fischer-Hübner i Rannenber 2011] podjął wiele problemów związanych z zarządzaniem tożsamością oraz zachowaniem prywatności w internecie, przedstawiając prototypy aplikacji z różnych obszarów, w tym uwierzytelniania czy budowania sieci społecznej. Jednym z większych projektów dotyczących tożsamości jest SWIFT [Rajasekaran i in. 2010], którego celem jest zbudowanie wielopoziomowych ram dla budowania aplikacji do zarządzania tożsamością. Istnieje wiele innych projektów podejmujących tematykę tożsamości, obszernie spisy takich projektów można znaleźć w internecie¹. Warto również zaznaczyć, że wokół problematyki zarządzania tożsamością koncentrują się prace wielu organizacji i inicjatyw, które koordynują i stymulują rozwój nowych rozwiązań. Jednymi z najważniejszych organizacji tego typu są: Identity Commons [2014], Kantara Initiative [2013] i OpenId Foundation [OpenID 2014].

¹ Na przykład: http://wiki.idcommons.net/Identity_Landscape, <http://blogs.law.harvard.edu/vrm/development/>.

2. System Ego

W poprzedniej części zaprezentowano przegląd literatury z zakresu wirtualnych tożsamości. W niniejszej sekcji omówiony zostanie system Ego oraz projekt o takiej samej nazwie, zajmujący się tym zagadnieniem, w ramach którego prowadzono badania przedstawione w prezentowanym artykule. Jak zdefiniowano w artykule [Węcowski, Małyszko i Abramowicz 2012], „celem projektu (Ego) jest opracowanie takiej struktury tożsamości, metod jej aktualizacji oraz modeli wymiany informacji między tożsamością i otoczeniem, które zaowocowałyby przeniesieniem miejsca składowania informacji o użytkowniku z serwerów dostawców usług do tożsamości użytkownika”. W trakcie powstawania projektu wypracowano m.in. szereg rozwiązań umożliwiających przetwarzanie tekstów w języku polskim w celu rozpoznawania tematyki danego tekstu oraz wykorzystania tak rozpoznanej tematyki artykułów do generowania semantycznych reprezentacji bytów związanych z danym tekstem.

Głównym obszarem zainteresowania badaczy w trakcie prac nad systemem Ego były tożsamości użytkowników. Za pomocą przygotowanego oprogramowania możliwe jest analizowanie artykułów czytanych przez użytkownika w przeglądarce internetowej (za jego zgodą) w celu pozyskiwania informacji o jego zainteresowaniach. Na tej podstawie możliwe jest generowanie jego tożsamości. W celu zapewnienia interoperacyjności tożsamości i jednocześnie jak największej wszechstronności informacji, którą taka tożsamość może zawierać, tożsamości generowane przez system wykorzystują strukturę hierarchii kategorii i haseł z Wikipedii z przypisanymi do nich wagami, co sprawia, że



Rysunek 1. Schemat procesu generowania wirtualnej tożsamości użytkownika w systemie Ego

tożsamość ma postać sieci semantycznej o ściśle zdefiniowanej i ogólnodostępnej strukturze. Dzięki temu dowolny system może przygotować moduł będący w stanie wykorzystać taką tożsamość.

Na rysunku 1 zaprezentowano schemat procesu, podczas którego generowana jest wirtualna tożsamość użytkownika. Specjalna wtyczka do przeglądarki monitoruje odwiedzane przez niego strony, z których następnie wybrana zostaje najważniejsza treść (dokument tekstowy). Kolejne kroki mają na celu wybór słów i fraz (tzw. słów kluczowych) w tekście, które potencjalnie mogą mieć znaczenie z punktu widzenia tematyki dokumentu. Takie słowa kluczowe są następnie ujednoznaczniane do conceptów z zasobu wiedzy (czyli haseł i kategorii z Wikipedii), z których to conceptów następnie wybierane są te, które prawdopodobnie najtrafniej reprezentują treść dokumentu. Taka reprezentacja dokumentu tekstowego może być wykorzystana do modyfikacji tożsamości użytkownika czytającego artykuł. Proces zmiany tożsamości na podstawie nowych informacji o użytkowniku (np. conceptów opisujących czytane przez niego artykuły) nazywamy aktualizacją tożsamości; proces ten został szczegółowo opisany w artykule [Węcowski i Schmidt 2014]. W tym procesie, w najprostszym ujęciu, wagi conceptów, które trafnie opisują obszar tematyczny danego artykułu, powinny w tożsamości wzrosnąć. Jest to umotywowane tym, że jeśli użytkownik czytał dany dokument, to prawdopodobnie interesuje się jego tematyką.

3. Stereotypy tożsamości i tożsamości bytów innych niż użytkownik

3.1. Tożsamości innych bytów

Dotychczas w niniejszym artykule zajmowaliśmy się wyłącznie tożsamościami użytkowników. Należy jednak zwrócić uwagę na to, że wypracowane rozwiązania mogą być zastosowane również na potrzeby generowania tożsamości innych bytów powiązanych z dokumentami tekstowymi. Przez powiązanie między bytem a artykułem rozumiemy, że byt wchodzi z danym artykułem w interakcję wskazującą na to, iż artykuł znajduje się w zakresie obszaru zainteresowania danego bytu. Przykładem takiej interakcji jest kliknięcie przez użytkownika linku do danego artykułu (gdzie bytem jest użytkownik), ale również umieszczenie przez portal danego artykułu na stronie głównej (gdzie bytem jest portal internetowy). Podobnie poprzez analizę artykułów umieszczanych w kanale RSS możliwe jest wygenerowanie tożsamości dane-

go kanału. Celem tego działania może być przykładowo stworzenie systemu służącego do wyszukiwania i rekomendacji takich kanałów.

Generowanie tożsamości dla dowolnych bytów może przebiegać w sposób analogiczny do generowania tożsamości użytkowników z tą różnicą, że w inny sposób pobierane są dokumenty powiązane z danym bytem. W przypadku portalu internetowego jako artykuły powiązane z profilowanym bytem można wykorzystać dokumenty pozyskane z tego portalu na drodze crawlingu. W przypadku kanału RSS dokumentami są zawartości stron, do których kierują hiperłącza znajdujące się w poszczególnych wiadomościach danego kanału. Jeśli posiada się zidentyfikowane artykuły powiązane z danym bytem, dalszy ciąg przetwarzania (ekstrakcja głównej zawartości strony, ekstrakcja słów kluczowych itd.) może być identyczny jak w przypadku procesu generowania tożsamości użytkownika, opisanym w części 2.

3.2. Generowanie stereotypów tożsamości użytkowników na podstawie tożsamości bytów z nim powiązanych

Jednym z podstawowych problemów napotykanym przez różne systemy informacyjne jest tzw. problem zimnego startu. Występuje on wówczas, gdy wymagane jest przeprowadzanie wnioskowania dla bytu, który dopiero niedawno pojawił się w systemie, przez co system nie dysponuje odpowiednią wiedzą konieczną dla tego celu. Przykładem takiego problemu może być np. rekomendowanie nowego produktu w bazie, który dotychczas nie został oceniony przez żadnego użytkownika [Schein i in. 2002]. Podobnie w przypadku interakcji z nowym użytkownikiem system początkowo nie posiada żadnej wiedzy pozwalającej na personalizację swojego działania pod kątem potrzeb użytkownika.

Rozwiązaniem tego problemu w przypadku użytkowników może być przypisanie im tzw. stereotypów. Stereotypy rozumiemy jako pewien początkowy zestaw cech przypisywanych użytkownikowi przez system na podstawie określonego wnioskowania. Pojęcie to po raz pierwszy zostało użyte w pracy [Rich 1979], w której na podstawie krótkiego dialogu z użytkownikiem system przypisywał mu zestaw stereotypów pochodzących z predefiniowanej puli. Przypisanie takich stereotypów pozwala od razu przeprowadzać wnioskowania o użytkownika, przy czym z oczywistych względów stereotyp nie jest równie dokładny jak profil nauczony na podstawie obserwacji zachowań użytkownika na przestrzeni dłuższego czasu. Z czasem jednak stereotyp może ewoluować tak, by lepiej odzwierciedlać zdobytą wiedzę o użytkowniku.

W projekcie Ego proponujemy generowanie stereotypów dla użytkowników na podstawie tożsamości bytów z nim powiązanych, przykładowo

portali informacyjnych oraz kanałów RSS. Scenariusz wygenerowania takiego stereotypu wyglądałby następująco. Nowy użytkownik systemu wprowadzałby adresy URL źródeł informacyjnych, które są zgodne z tematyką jego zainteresowań (przykładowo kanałów RSS, które subskrybuje). Artykuły udostępniane w tych źródłach byłyby następnie przetwarzane i w rezultacie generowana byłaby tożsamość, którą można przypisać do użytkownika jako jego tożsamość początkową, służącą jako źródło informacji o użytkowniku, jeszcze zanim system zdoła zebrać o nim bardziej szczegółowe informacje. Taka początkowa tożsamość podlegałaby dalszej ewolucji w standardowy sposób, na podstawie analizy stron czytanych przez użytkownika.

4. Opis eksperymentu

4.1. Cel eksperymentu

Celem eksperymentu była odpowiedź na pytanie: Czy tożsamości źródeł informacyjnych wygenerowane zgodnie z rozwiązaniem wypracowanym w trakcie tworzenia projektu dobrze odzwierciedlają tematykę artykułów w danym źródle? Teza, którą starano się udowodnić, brzmiała, że tożsamość źródła trafnie opisuje tematykę publikowanych w nim artykułów. Aby tego dowieść, porównywano podobieństwo wygenerowanej tożsamości źródła do artykułów w nim opublikowanych oraz do artykułów pochodzących z innych źródeł. Brak podobieństwa tożsamości źródła do artykułów w nim publikowanych oznaczałoby, że tożsamości źródeł informacyjnych nie pozwalają na wnioskowanie o tematyce artykułów w nich umieszczanych, a tym samym o zainteresowaniach użytkowników korzystających z danego źródła.

4.2. Dane

Eksperyment przeprowadzono na kolekcji 17 kanałów RSS o różnej tematyce. Spośród nich wybrano cztery kanały, dla których wygenerowano tożsamości. Były to:

- 1) deon – kanał religijny, tożsamość skonstruowano na podstawie 55 dokumentów,
- 2) gazeta.pl – kanał sportowy, tożsamość skonstruowano na podstawie 150 dokumentów,
- 3) interia.pl – kanał motoryzacyjny, tożsamość skonstruowano na podstawie 96 dokumentów,
- 4) gazeta.pl – kanał nowoczesne technologie, tożsamość skonstruowano na podstawie 60 dokumentów.

Tabela 1. Najważniejsze koncepty dla tożsamości dwóch wybranych kanałów RSS

Deon – kanał religijny		Interia – kanał motoryzacyjny	
koncept	waga	koncept	waga
chrześcijaństwo	0,199	pojazdy	0,165
religie_abrahamiczne	0,119	matematyka	0,094
urzędy_i_tytuły_kościelne	0,109	motoryzacja	0,088
religioznawstwo	0,089	transport	0,083
kult_religijny	0,085	samochód	0,083
papież	0,075	kierowca	0,077

Źródło: na podstawie metody opisanej w artykule [Węcowski i Schmidt 2014].

Dla tych kanałów wygenerowano tożsamości z wykorzystaniem modułów opisanych w artykule [Węcowski i Schmidt 2014]. Metoda ta, otrzymując na wejściu reprezentacje dokumentów w postaci powiązanych z nimi konceptów z Wikipedii, w odpowiedni sposób modyfikuje wagi przypisane do tych konceptów w tożsamości, obrazując w ten sposób zyskiwaną informację o tym, że użytkownik interesuje się danym zagadnieniem². Przykłady wygenerowanych tożsamości zaprezentowano w tabeli 1. Dodatkowo w kolekcji znajdowało się 1015 artykułów ze wspomnianych 17 źródeł (odrzucono artykuły wykorzystane do wygenerowania tożsamości).

4.3. Przebieg eksperymentu

Każdą z czterech wygenerowanych tożsamości porównano kolejno ze wszystkimi posiadanymi artykułami. Dla każdej tożsamości część porównywanych z nią artykułów pochodziła z kanału, dla którego wygenerowano daną tożsamość, a część (przeważająca) z pozostałych 16 kanałów. Porównywanie artykułu z tożsamością polegało na tym, że obliczano podobieństwo zgodnie z miarą cosinusa pomiędzy tożsamością a reprezentacją dokumentu mającą postać par (*koncept z Wikipedii, waga*), wygenerowaną zgodnie z metodą opisaną w części 2. W związku z tym dla każdej tożsamości otrzymywano dwie populacje:

- podobieństwa tożsamości źródła do artykułów ze źródła, dla którego tożsamość była wygenerowana (X),
- podobieństwa tożsamości źródła do artykułów z innych źródeł (Y).

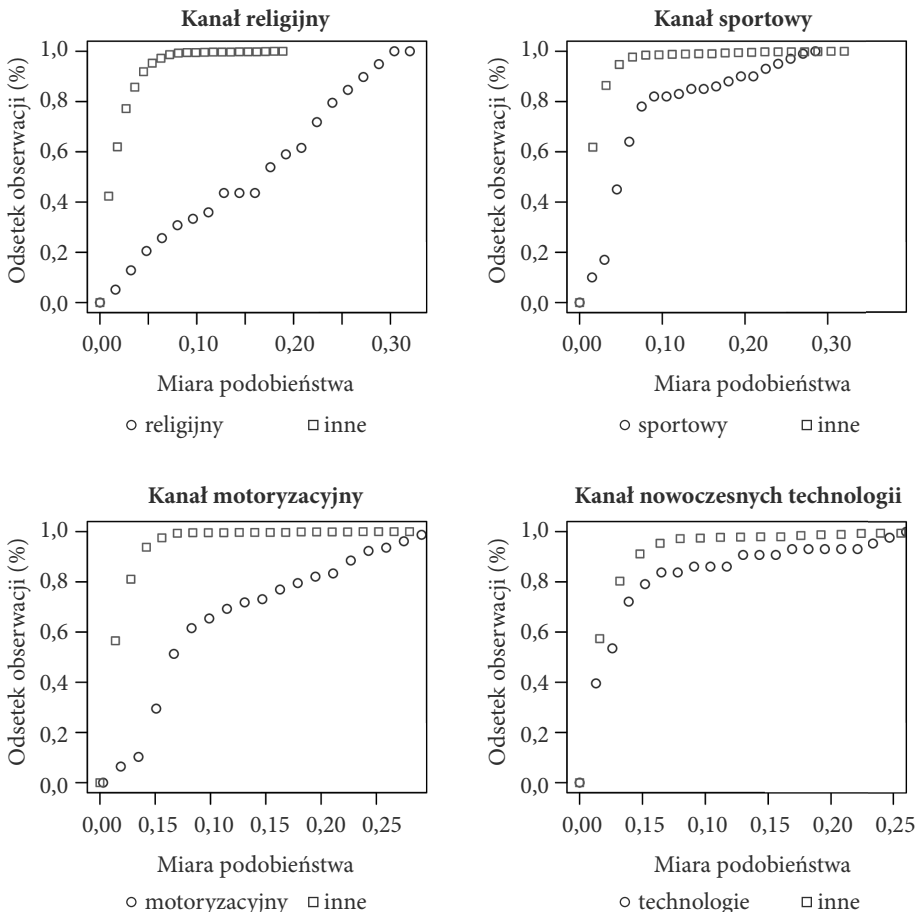
² Bardziej szczegółowy opis tej metody wymagałby wiele miejsca i wykraczałby poza ramy tego artykułu; zainteresowanego czytelnika odsyłamy do wspomnianego artykułu, który jest dostępny w internecie.

Następnie porównano obie populacje, aby ustalić, czy mają jednakowy rozkład. Porównanie rozkładów populacji przeprowadzono za pomocą testu Manna-Whitneya, w którym przyjęto następujące hipotezy:

- populacje X i Y mają ten sam rozkład podobieństwa,
- rozkłady w obu populacjach są różne.

4.4. Wyniki

Na rysunku 2 zaprezentowano wykresy skumulowanych rozkładów podobieństw dla obu populacji (X i Y) dla każdej tożsamości. W każdym przypadku wykres skumulowanego rozkładu dla podobieństwa tożsamości do artykułów



Rysunek 2. Dystrybuanta podobieństwa obserwacji do tożsamości poszczególnych kanałów

z innych źródeł jest przesunięty w lewo i w górę. Oznacza to, że w każdym przypadku większy odsetek obserwacji z tej populacji ma podobieństwo na niskim poziomie. Najbardziej jest to widoczne w przypadku tożsamości źródła deon.pl, gdzie 80% artykułów z innych źródeł ma podobieństwo niższe niż 0,05, podczas gdy dla artykułów z tego źródła odsetek ten wynosi zaledwie około 20%. W pozostałych przypadkach ta różnica w skumulowanych rozkładach nie jest już tak wielka, jednak ciągle jest wyraźna (przy czym najmniejsza jest dla kanału nowoczesnych technologii).

Przeprowadzony test Manna-Whitneya potwierdził te obserwacje. W trzech przypadkach na cztery pozwolił on na odrzucenie hipotezy zerowej przy poziomie ufności równym 5%; nie udało się jej odrzucić wyłącznie w przypadku kanału nowoczesnych technologii. Jak widać na rysunku, również na tym kanale widoczna jest różnica w obu populacjach, jest ona jednak zbyt niska, aby odrzucić hipotezę zerową przy określonym poziomie ufności.

Na podstawie przeprowadzonego wnioskowania statystycznego należy stwierdzić, że tożsamości wygenerowane dla źródeł informacyjnych zgodnie z opisanym podejściem trafnie reprezentują tematykę artykułów zamieszczanych w tych źródłach. Tym samym przypisanie takiej tożsamości do użytkownika deklarującego zainteresowanie danym portalem może przynieść informacje użyteczne dla przeprowadzania wnioskowań o zainteresowaniach użytkownika.

Zakończenie

Przedstawiona praca podejmuje tematykę związaną z profilowaniem użytkowników internetu oraz innych bytów za pomocą wirtualnych tożsamości. W pracy zaproponowano podejście do dwóch problemów występujących we wspomnianym obszarze – problemu budowania tożsamości początkowej przy ograniczonej informacji o profilowanym bycie oraz problemu przeniesienia koncepcji tożsamości osoby na byty inne niż człowiek. Opisano metodę budowania stereotypu tożsamości jako możliwego rozwiązania problemu „zimnego startu” przy budowie tożsamości. Przedstawiono również podejście wykorzystujące tożsamości do profilowania źródeł informacyjnych dostępnych w internecie.

Zaprojektowany i pozytywnie ewaluowany eksperyment potwierdził przydatność wirtualnych tożsamości w opisie źródeł informacyjnych. Wskazał również, że tożsamości bytów manifestujących się w internecie mogą posłużyć do budowania stereotypów tożsamości dla nowych użytkowników.

Bibliografia

- Camenisch, J., Fischer-Hübner, S., Rannenberg, K. (ed.), 2011, *Privacy and Identity Management for Life*, Springer, Berlin Heidelberg.
- Cameron, K., 2005, *The Laws of Identity*, <http://msdn.microsoft.com/en-us/library/ms996456.aspx> [dostęp 14.08.2013].
- Clarke, R., *Dissidentity*, 2008, w: redaktor *Identity in the Information Society*, Springer, <https://link.springer.com/article/10.1007%2Fs12394-009-0013-7> [dostęp 11.12.2017].
- Gilbert, C., Upatising, L., 2013, *Formal Analysis of BrowserID/Mozilla Persona*, <http://www.stanford.edu/~lazau/BrowserIDPersona.pdf> [dostęp 20.11.2013].
- Hammer-Lahav, E., 2011, *The OAuth 1.0 Protocol*, <http://tools.ietf.org/html/rfc5849> [dostęp 21.11.2013].
- Higgins, 2013, *PDS Vision*. http://wiki.eclipse.org/PDS_Vision [dostęp 14.11.2013].
- Hodges, J., Philpott, R., Maler, E., 2005, *Glossary for the OASIS Security Assertion Markup Language (SAML) V2.0*, <http://docs.oasis-open.org/security/saml/v2.0/saml-glossary-2.0-os.pdf> [dostęp 5.11.2013].
- Identity Commons, 2013, <http://www.idcommons.org/> [dostęp 12.08.2013].
- Identity Commons. Identity Landscape, 2012, http://wiki.idcommons.net/Identity_Landscape [dostęp 12.08.2013].
- Kantara Initiative, 2013, <http://kantarainitiative.org/> [dostęp 12.08.2013].
- Kubicek, H., 2010, *Introduction: Conceptual Framework and Research Design for a Comparative Analysis of National eID Management Systems in Selected European Countries*, *Identity in the Information Society*, vol. 3, iss. 1, s. 5–26, Springer, Berlin Heidelberg.
- Miorandi, M. Sicari, S., De Pellegrini, F. Chlamtac, I., 2012, *Internet of Things: Vision, Applications and Research Challenges*, *Ad Hoc Networks*, vol. 10, iss. 7, s. 1497–1516.
- Nabeth, T., 2009, *Identity of Identity*, in: Rannenberg, K., Royer, D., Deuker, A. (eds.), *The Future of Identity in the Information Society: Challenges and Opportunities*, Springer, Berlin Heidelberg, s. 19–69.
- OpenID Community, *OpenID Authentication 2.0 – Final*, 2007, http://openid.net/specs/openid-authentication-2_0.html [dostęp 20.11.2013].
- OpenID Foundation. Work Groups of OpenID Foundation, 2013, <http://openid.net/wg/> [dostęp 12.08.2013].
- Personal Inc., 2013, *Personal system*, <https://www.personal.com/> [dostęp 12.08.2013].
- Ragouzis, N., Hughes, J., Philpott, R., Maler, E., 2006, *Security Assertion Markup Language (SAML) V2.0 Technical Overview*, http://www.oasis-open.org/committees/documents.php?wg_abbrev=security [dostęp 20.11.2013].
- Rajasekaran, H. et al., 2010, *SWIFT White Paper. SWIFT Identity Architecture*, http://www-wordpress.sit.fraunhofer.de/ist-swift/wp-content/uploads/sites/10/2013/10/Whitepaper-SWIFT_Identity_Architecture.pdf [dostęp 14.11.2013].

- Rannenber, K., Royer, D., Deuker, A. (eds.), 2009, *The Future of Identity in the Information Society: Challenges and Opportunities*, Springer, Berlin Heidelberg.
- Rich, E., *User Modeling via Stereotypes*, 1979, *Cognitive Science*, vol. 3, iss. 4, s. 329–354.
- Rundle, M., Maler, E., Nadalin, A., Reed, D., Thibeau, D., 2010, *The Open Identity Trust Framework Model*, <http://openididentityexchange.org/sites/default/files/the-open-identity-trust-framework-model-2010-03.pdf> [dostęp 14.08.2013].
- Scerri, S., Gimenez, R., Herman, F., Bourimi, M., Thiel, S., 2011, *Digital.Me towards an Integrated Personal Information Sphere*, in: *Proceedings on the Federated Social Web Summit, W3C*.
- Schein, A., Popescul, A., Ungar, L.H., Pennock, D.M., 2002, *Methods and Metrics for Cold-start Recommendations*, *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, s. 253–260.
- Story, H., Corlosquet, S., 2011, *Web 1.0. Web Identification and Discovery*, <http://www.w3.org/2005/Incubator/webid/spec> [dostęp 20.11.2013].
- Węcowski, D.G., Małyszko, J., Abramowicz, W., 2012, *Zarządzanie tożsamością w internecie przyszłości*, w: Jurek, W. (red.), *Matematyka i informatyka na usługach ekonomii. Informatyka ekonomiczna*, Zeszyty Naukowe, nr 239, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, s. 9–19.
- Węcowski, D.G., Schmidt P., 2014, *Evolution of Wikipedia-based Virtual Identities*, *Journal of Systems Integration*, vol. 5, iss. 3, s. 67–77.