

Ilona Pawełoszek-Korek

NIEJAWNE SPOŁECZNOŚCI WIRTUALNE JAKO ŹRÓDŁO WIEDZY PRZEDSIĘBIORSTWA

1. Wstęp

Coraz większa liczba przedsiębiorstw przenosi swoje działania gospodarcze na arenę wirtualną. W gospodarce sieciowej załamuje się wiele paradygmatów klasycznej ekonomii, przede wszystkim pojawia się w niej dominacja zasobów niematerialnych (kapitału intelektualnego, emocjonalnego, wiedzy itp.) nad kapitałem materialnym i finansowym (Perechuda 2005, s. 117).

Istnieje wiele organizacji, które kierują swoje strategie marketingowe głównie, bądź wyłącznie, na rynki elektroniczne, dają one bowiem niemal nieograniczone możliwości pozyskania klientów, a koszty prowadzonej w ten sposób działalności są nieporównywalnie mniejsze niż koszty tradycyjnych kanałów dystrybucji. Oczywiście poszerzenie rynku zbytu oznacza również zwiększenie konkurencji, ponieważ firma działająca w Internecie może rozpatrywać jako swoich rywali wszystkie przedsiębiorstwa określonej branży, których strony www znajdują się wyżej w rankingu wyszukiwarek lub są na zbliżonej pozycji.

Sieć Internet w obecnych czasach stała się nie tylko medium przesyłania danych na odległość, ale coraz częściej traktowana jest, zarówno przez klientów, jak i przedsiębiorstwa, jako źródło wiedzy. Grupy osób korzystających z Internetu tworzą struktury społeczne łączące ludzi określoną więzią (np. zainteresowanych tym samym produktem, korzystających z podobnych usług różnych dostawców, bądź jednego dostawcy, czy też poszukujących informacji na ten sam temat).

Zmiana sposobu działalności z tradycyjnej na elektroniczną oraz rynku z lokalnego na globalny niesie ze sobą konieczność zastosowania nowych technik zbierania informacji, pozyskiwania wiedzy o rynku, produktach i klientach na potrzeby skutecznego zarządzania. W tym celu można skorzystać z technik jest odkrywania i badania społeczności webowych tworzących się wokół danej firmy i jej konkurentów. Techniki, o których mowa, mogą mieć również zastosowanie do badania i wzbogacania wiedzy niezbędnej w funkcjonowaniu przedsiębiorstw tradycyjnych,

ponieważ sieć Internet często stanowi forum wymiany poglądów i informacji dotyczących różnorodnych tematów. Celem niniejszego artykułu jest prezentacja istoty niejawnych społeczności wirtualnych w kontekście wykorzystania ich jako cennego źródła wiedzy dla przedsiębiorstw.

2. Istota i źródła powstawania społeczności wirtualnych

Społeczność stworzona dzięki wykorzystaniu globalnej sieci komputerowej jest określana jako wirtualna, webowa, internetowa, elektroniczna. Nazwy te stanowią synonimy. W tym artykule używane jest określenie „społeczność wirtualna”. Wirtualność określa się przez charakterystykę własności, a nie istniejące cechy fizyczne (Kisielnicki, Sroka 2005, s. 226). Słowo to pochodzi z łaciny i oznacza byt, który może wywoływać skutki fizyczne, nie będąc bytem fizycznym, czyli rozpoznawalnym za pomocą zmysłów. Społeczność wirtualną „tworzą osoby prywatne lub przedsiębiorstwa czy inne organizacje o wspólnych zainteresowaniach, które komunikują się za pośrednictwem Internetu” (Sznajder 2002, s. 138).

Castells określa społeczność wirtualną jako samodefiniującą się elektroniczną sieć interaktywnej komunikacji pomiędzy jej użytkownikami, zorganizowaną wokół wspólnych zainteresowań lub celu, czasami celem jest sama komunikacja (Castells 1996, s. 362). Uczestnicy społeczności wirtualnej korzystają z różnych zasobów informacyjnych w sieci, jednocześnie wzbogacają je przez dodawanie wartości. Na przykład tworzą odnośniki do zasobów, opinie, komentarze, porównania, uzupełnienia. Uczestnicy społeczności wirtualnej są więc pośrednio zaangażowani w osiąganie celów przedsiębiorstwa (zależnie od kierunku swojej aktywności i doświadczeń mogą ułatwiać lub utrudniać firmie ich osiągnięcie). Uczestnicy mogą być sklasyfikowani jako wewnętrzni (np. klienci czy też pracownicy przedsiębiorstwa) oraz zewnętrzni (konkurencja, potencjalni klienci).

Przedsiębiorstwa często same inicjują powstawanie społeczności sieciowych, których celem jest szerzenie wiedzy o produktach wśród klientów. Wtedy można mówić o społecznościach jawnych, czyli takich, które łatwo zidentyfikować, odnaleźć i analizować. Narzędziami służącymi tworzeniu społeczności mogą być fora, listy, dyskusyjne, czaty. Przykładami dobrze rozwiniętych jawnych społeczności mogą być np. strony klubu Nokia, zrzeszające użytkowników urządzeń mobilnych. Społeczności określane jako niejawne natomiast powstają spontanicznie (nie są kierowane ani inicjowane przez zainteresowane organizacje).

3. Niejawne społeczności webowe jako źródło wiedzy

Strony firm, wyposażone w narzędzia wymiany myśli i opinii, takie jak fora dyskusyjne, są łatwe do odnalezienia i zbadania, ponieważ ich struktura jest z góry ustalona i przeważnie podzielona na kategorie tematyczne. Fora dyskusyjne czy też listy dyskusyjne prowadzone przez daną firmę zapewniają zatem łatwy dostęp do wiedzy. Informacje pochodzące z jednego źródła mogą być gromadzone w bazie

danych, archiwizowane i mogą stanowić materiał do analiz. Dzięki wypowiedziom na forum można ustalić m.in., jakie są oczekiwania klientów wobec firmy, jakie problemy najczęściej się pojawiają podczas korzystania z produktów, jakie rozterki towarzyszą potencjalnym klientom, jak produkty są postrzegane na tle konkurencji, czy dobrze jest oceniana jakość obsługi itp.

Strony www danej firmy są powszechnie dostępne w Internecie i mogą stanowić także źródło wiedzy dla konkurencji, dlatego też kierownictwo firmy może przeciwdziałać umieszczaniu informacji mogących w jakikolwiek sposób narazić *image* przedsiębiorstwa lub przynieść straty. Fora i listy dyskusyjne mogą być moderowane i cenzurowane przez ich administratorów. Zatem analiza forum dyskusyjnego lub zapisanie się na listę dyskusyjną, aby zdobyć informacje o konkurencyjnej firmie, nie zawsze muszą dawać rzeczywisty obraz konkurencji. Wielokrotnie ciekawe informacje o otoczeniu firmy, własnej bądź konkurencyjnej, mogą być uzyskane przez odkrywanie i analizę ukrytych społeczności danej organizacji.

Ukryte społeczności mają tę zaletę, iż trudniej nimi sterować, a zatem informacje uzyskane za ich pomocą mogą być bardziej miarodajne. Jednakże zawsze należy uwzględnić istnienie szumów informacyjnych i starać się je odróżnić od cennych informacji.

Identyfikacja ukrytych społeczności daje firmie sygnał, jak daleko sięga jej oddziaływanie. Można się dowiedzieć, z jakich krajów pochodzą uczestnicy społeczności, kim są, jakimi językami głównie się posługują, jakie firmy stanowią konkurencję na oddalonych geograficznie rynkach lokalnych.

4. Wykrywanie niejawnych społeczności wirtualnych

Społeczność wirtualną można zbadać jedynie przez analizę źródeł elektronicznych, takich jak strony internetowe, listy dyskusyjne, wiadomości poczty elektronicznej lub komunikatorów internetowych. Dlatego z punktu widzenia badawczego może być ona traktowana jako grupa obiektów (przeważnie stron www), które mają jakąś wspólną cechę, np. odnoszą się do tego samego tematu, są założone przez ludzi o określonych zainteresowaniach, mają hiperłącza do tych samych stron. Oczywiście z punktu widzenia socjologicznego społeczność tworzą ludzie, a strony internetowe są tylko artefaktami mówiącymi o istnieniu więzi międzyludzkiej.

Społeczności wirtualne mają zatem swoją strukturę semantyczną, w ramach której można wyróżnić mniejsze zbiory (np. społeczność zainteresowana urządzeniami mobilnymi zrzesza osoby, wśród których można wyróżnić społeczności użytkowników, programistów, projektantów, handlowców czy też społeczności zainteresowane tylko konkretnym modelem urządzenia).

Jak wynika z dotychczasowych rozważań, społeczności wirtualne stanowią skomplikowane sieci powiązań, na które można patrzeć wielowymiarowo. Strony

www stanowią obiekty połączone różnymi, niewidocznymi na pierwszy rzut oka zależnościami, które można odkryć dopiero po dokonaniu szczegółowej analizy statystycznej. Dlatego też do ich badania dobrze nadają się techniki *data mining*.

Do zbierania danych na temat powiązań stron można wykorzystać popularne wyszukiwarki internetowe, takie jak Google, AltaVista, HotBot, Excite czy Yahoo, które oferują bogate możliwości zadawania zapytań. Wymienione wyszukiwarki pozwalają na odnalezienie stron podobnych do podanej oraz zawierających hiperłącza do wskazanej witryny.

Do prowadzenia regularnych badań dobrze jest skonstruować własną aplikację, dostosowaną do konkretnych potrzeb organizacji, korzystającą z wyników kilku wyszukiwarek. Wiele spośród nowoczesnych wyszukiwarek oferuje tzw. WebServices, czyli usługi pozwalające zadać zapytanie i pobrać wyniki wyszukiwania w postaci pliku XML. Takie rozwiązanie, ze względu na uniwersalność języka XML, znacznie ułatwia tworzenie aplikacji korzystających z wielu różnych źródeł informacyjnych.

5. Metody badania społeczności wirtualnych

Analiza sieci społecznych (ang. *social network analysis*) polega na mapowaniu i mierzeniu relacji oraz przepływów między ludźmi, grupami, organizacjami, komputerami lub innymi obiektami przetwarzającymi informacje/wiedzę (Cai i in. 2005, s. 59]. Metody badania społeczności wirtualnych wywodzą się z dorobku analizy sieci społecznych.

Techniki wykrywania powiązań między obiektami są znane od dawna w nauce bibliometrii, która zajmuje się m.in. badaniem dorobku naukowego w różnych dziedzinach. Artykuły naukowe są analizowane pod kątem liczby cytowań w innych publikacjach. W ten sposób tworzony jest indeks cytowań. W sieci istnieje wiele takich indeksów publikacji naukowych, z których najbardziej znane to: Google Scholar, MathSciNet, Science Citation Index. Analizowane w ten sposób dokumenty obejmują publikacje akademickie, recenzje, reprinty, książki, abstrakty i raporty techniczne z różnych dziedzin badań. Bazy te zawierają informacje o organizacjach związanych z autorami, więc są często używane do analizy konkurencyjnej wśród wydawnictw.

Borgman i Furner (za: Reid 2003) definiują dwa rodzaje analizy hiperłączy: relacyjną i oceniającą. Relacyjna analiza daje wgląd w siłę powiązań między obiektami na określonych stronach, natomiast analiza oceniająca mówi o popularności lub jakości czy też istotności strony www dla badanego obszaru.

6. Ocena zawartości strony www

Statystyczną analizę stron internetowych pod kątem ich istotności dla danej dziedziny (zapytania użytkownika) dokonuje się m.in. na postawie ustalenia liczby

cytowań (odnośników, hiperłączy) do danego tekstu. Dany obiekt w sieci jest uznawany za tym ważniejszy, im więcej stron odwołuje się do niego.

Ważność strony internetowej określana jest według określonych algorytmów, z których najbardziej znane to HITS i PageRank.

Algorytm HITS został opracowany w 1998 r. przez Kleinberga, który zaproponował definicję istotności strony www. Uważa, że nie jest istotne, aby wartościowe strony (w języku angielskim określane jako *authorities*) wskazywały za pomocą hiperłączy inne wartościowe strony. Zamiast tego istnieją specjalne węzły w sieci określane jako „huby”, które zawierają zbiór linków do dobrych stron. Zaproponował dwupoziomowy schemat propagacji wag, gdzie poparcie jest udzielane dobrym stronom poprzez „huby” częściej niż bezpośrednio pomiędzy właściwymi stronami. W tym modelu każda strona może być rozpatrywana w dwojaki sposób: jako „hub”, który wskazuje na wartościowe zasoby, i jako wartościowa strona, zawierająca sama cenne informacje. Zatem istnieje wzajemna relacja wzmacniająca pozycję stron. Dobry „hub” jest stroną, która wskazuje inne wartościowe strony, natomiast wartościowa strona jest tą, którą wskazują dobre „huby” (Borodin i in. 2005, s. 235). Na tej podstawie można obliczać wagowe wartości stron www.

PageRank jest algorytmem zakładającym, że linki z wartościowych (dobrych jakościowo) stron powinny być traktowane jako bardziej cenne niż inne linki. Ważne jest to, aby wiedzieć, jak wiele stron odsyła do danej strony, a także to, czy jakość tych stron jest wysoka. Algorytm PageRank tworzy schemat jednopoziomowej propagacji wag, gdzie dobrą stroną jest ta, która jest wskazywana przez wiele dobrych stron (Borodin i in. 2005, s. 234).

7. Przykłady badań wirtualnych społeczności niejawnych

Ciekawe rozwiązanie badania i wykrywania niejawnych społeczności wirtualnych proponuje E. Reid (2003). Procedura badawcza oparta na koncepcji analizy hiperłączy składa się z następujących kroków:

- 1) identyfikacji docelowego przedsiębiorstwa i adresów URL jego stron www do badania,
- 2) stworzenia „mapy otoczenia” identyfikującej uczestników społeczności przedsiębiorstwa, takich jak klienci, dostawcy itp.
- 3) przeprowadzenia przeszukiwania hiperłączy na podstawie danego adresu URL,
- 4) zakodowania zebranych danych,
- 5) identyfikacji i analizy wzorców.

Procedura została praktycznie wykorzystana do identyfikacji społeczności wirtualnych zgromadzonych wokół dużej firmy dostarczającej systemów informatycznych „MicroStrategy”. Do przeprowadzenia badania hiperlinków wchodzących do strony firmy zostały wykorzystane zaawansowane opcje wyszukiwarki Google.

W ten sposób można ustalić jedynie przybliżoną liczbę linków wchodzących do strony, ponieważ niektóre źródła odwołujące się do niej są zabezpieczone przed indeksowaniem za pomocą *firewalls* lub innych podobnych technik.

Kodowanie zebranych danych polega na usunięciu zduplikowanych hiperłączy i związanych z nimi tekstów oraz samocytujących się źródeł (linków w obrębie witryny firmy do jej podstron). Dane zostały przejrane pod kątem wyodrębnienia grup uczestników społeczności wirtualnych, zgodnych z ustaloną „mapą otoczenia”. Uczestnicy zostali podzieleni na wewnętrznych (partnerzy/dostawcy, klienci, pracownicy, inwestorzy) i zewnętrznych (instytucje oświatowe, media, portale, osoby publiczne, konkurenci, kandydaci na stanowiska w firmie i in.). Dokonano klasyfikacji danych według krajów z których pochodziły, domen internetowych, instytucji, języków, w jakich pojawiały się teksty.

Dzięki badaniu ustalono, że społeczności niejawne są zainteresowane także takimi zasobami niematerialnymi firmy, jak systemy *Business Intelligence*, kreowanie marki, strategia. Ustalono także, w jaki sposób uczestnicy niejawnych społeczności wirtualnych korzystają z zasobów udostępnianych na stronach internetowych firmy. Kiedy tworzą hiperłącza do stron firmy, często umieszczają obok nich swoje opinie, komentarze, spostrzeżenia. Systematyczne zbieranie i analiza takich informacji – poprzednio nie branych pod uwagę – mogą być źródłem wiedzy dla przedsiębiorstwa.

Opinie i odczucia, zwłaszcza jeśli są odmienne od powszechnie znanych, mogą być cennym źródłem dla systemów wczesnego ostrzegania. Badanie niejawnych społeczności internetowych można zatem uznać za działanie wspomagające procesy *Business Intelligence*.

Dotychczasowe badania dotyczące bibliotek cyfrowych i ich społeczności sugerowały automatyczne podejście do zbierania stron www w określonych dziedzinach.

Jednakże gromadzenie stron www do badań jest procesem zbierania i organizowania niestukturalnych informacji, dlatego nie zawsze jest możliwe w pełni automatyczne przeprowadzenie badania.

Studium wykorzystania technik badania społeczności niejawnych prezentują E. Reid i in. (2007). W studium tym wykorzystano webspidera do wykrywania społeczności niejawnych terrorystów islamskich. Strony terrorystów posłużyły jako materiał wyjściowy do przeprowadzenia wyszukiwania hiperłączy zwrotnych (korzystano z zaawansowanego wyszukiwania Google). W badaniu tym nie zautomatyzowano w pełni procesu przeszukiwania, zastępując ten mechanizm oceną dokonaną przez ludzkich ekspertów. Podejście to podyktowane było tym, iż w sieci znajduje się bardzo dużo stron fałszywych mówiących o terroryzmie, ale nie będących wytworem grup terrorystycznych, które były przedmiotem badania.

Metody badania społeczności sieciowych przeważnie wykorzystują hiperłącza pomiędzy stronami, jednakże istnieje również inne, warte uwagi, podejście do wykrywania podobieństw między dokumentami zawierającymi tekst. X. Li i B. Liu

(2005, s. 275-276) opisują metodę wykrywania struktur sieciowych z tekstu. Analiza zbioru dokumentów pod kątem występowania określonych słów kluczowych prowadzi do wyodrębnienia zdań dotyczących określonej encji (obiektu, tematu). Wyselekcjonowane zdania są następnie mapowane w postaci ważonego grafu. Autorzy proponują hierarchiczny algorytm klastrujący do analizy grafów pozwalający odkryć w nich społeczności.

Do przeprowadzenia przeszukiwania stron www pod kątem występowania słów kluczowych można użyć dowolnej wyszukiwarki, bez potrzeby korzystania z opcji zaawansowanych. Zakładając, iż nie wszystkie strony www poruszające określoną tematykę muszą zawierać hiperłącza do innych podobnych zasobów, można stwierdzić, że mówione metody (analiza hiperłączy oraz wykrywanie struktur sieciowych z tekstu) mogą się wzajemnie dobrze uzupełniać.

Większość algorytmów używanych do badania powiązań w społecznościach uwzględnia tylko jedną sieć powiązań, w której występuje pojedyncza relacja (np. strony połączone hiperłączami). W rzeczywistości, w społecznościach zawsze istnieje wiele relacji, a każda z nich może być traktowana jako sieć relacji. Taki rodzaj powiązań może być określany jako heterogeniczna sieć powiązań. Aby odkryć istnienie społeczności mającej określone cechy, najpierw należy ustalić, jakie relacje odgrywają ważną rolę w takiej społeczności. Należy podkreślić, że takie relacje nie zawsze są jawne. (Cai i in. 2005, s. 58).

8. Podsumowanie

Sieć Internet stwarza bogate możliwości badania struktur społecznych, dzięki zaawansowanym mechanizmom wyszukiwania. Informacje pozyskane z badań mogą posłużyć przedsiębiorstwom jako źródło wiedzy nie tylko o konkurencji, klientach czy tendencjach na rynku, ale także o nich samych. Niejawne społeczności sieciowe mogą dostarczyć danych dla systemów wczesnego ostrzegania i rozwiązań typu *Business Intelligence*. Artykuł ten stanowi wprowadzenie do tematyki badania niejawnych społeczności wirtualnych. Dziedzina ta, z powodu rozwoju sieci Internet i ciągłego zwiększania liczby stron www o różnorodnej zawartości, wymaga wypracowania nowych, skutecznych podejść badawczych, narzędzi pozwalających ułatwić zarówno zbieranie danych, jak i ich analizę pod kątem uzyskania cennych dla przedsiębiorstw informacji.

Literatura

- Borodin A., Roberts G.O., Rosenthal J.S., Tsaparas P., *Link Analysis Ranking: Algorithms, Theory, and Experiments*, „ACM Transactions on Internet Technology” 2005, 5(1).
Cai D., Shao Z., He X., Yan X., Han J., *Mining Hidden Community in Heterogeneous Social Networks*, [w:] *Proceedings of the 3rd international Workshop on Link Discovery*, Chicago 2005.

- Castells M., *The Rise of the Network Society*, Blackwell, Malden 1996.
- Kisielnicki J., Sroka H., *Systemy informacyjne biznesu*, Placet, Warszawa 2005.
- Li X., Liu B., *Mining Community Structure of Named Entities from Free Text*, [w:] *Proceedings of the 14th ACM international Conference on information and Knowledge Management, Bremen, Germany*, ACM Press, New York 2005.
- Perechuda K., *Dyfuzja wiedzy w przedsiębiorstwie sieciowym. Wizualizacja i kompozycja*, AE, Wrocław 2005.
- Reid E., *Using Web Link Analysis to Detect and Analyze Hidden Web Communities*, [w:] *Information and Communications Technology for Competitive Intelligence*, red. D. Vriens, Irm Press 2003, s. 57-84.
- Reid E., Qin J., Zhou Y., Lai G., Sageman M., Weimann G., Chen H., *Collecting and Analyzing the Presence of Terrorists on the Web: A Case Study of Jihad Websites*, źródło internetowe: http://ai.arizona.edu/research/terror/publications/ISI_AILab_submission_final.pdf z dnia 14.02.2007.
- Sznajder A., *Marketing wirtualny*, Oficyna Ekonomiczna, Kraków 2002.
- Zhang Y., Yu J.X., Hou J., *Web Communities Analysis and Construction*, Springer-Verlag, Berlin Heidelberg 2006.

HIDDEN WEB COMMUNITIES AS SOURCES OF KNOWLEDGE FOR AN ENTERPRISE

Summary

This paper addresses the detection and analysis of hidden Web communities based on information available from Web, such as Web document content and hyperlinks. Web communities are interesting research domain and they can become a valuable resource of knowledge for business intelligence systems of an enterprise. Applications of aforementioned methods are another aspect emphasized in this paper. Before presenting various approaches to detect and analyze hidden web communities, some definitions of hidden web communities are provided according to different authors.