Uniwersytet Ekonomiczny w Katowicach

Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 355 · 2018

Informatyka i Ekonometria 13

Wiesław Wolny

Uniwersytet Ekonomiczny w Katowicach
Wydział Informatyki i Komunikacji
Katedra Informatyki
wieslaw.wolny@uekat.pl

BEZPIECZEŃSTWO I PRYWATNOŚĆ DANYCH W BADANIACH MEDIÓW SPOŁECZNOŚCIOWYCH

Streszczenie: Analiza mediów społecznościowych pozwala na ujawnienie informacji nie tylko wprost opublikowanych na portalach, ale również tych, które zostały odkryte różnymi metodami. Publikacja takich danych może stanowić naruszenie prywatności. Artykuł przedstawia możliwe problemy badawcze w tym zakresie. Celem artykułu jest zarysowanie koncepcji rozwiązań zapewniających bezpieczeństwo i prywatność w badaniach mediów społecznościowych.

Słowa kluczowe: media społecznościowe, odkrywanie wiedzy, ochrona danych osobowych.

JEL Classification: C10, C18.

Wprowadzenie

Ochrona prywatności jest jednym z najważniejszych wyzwań mediów społecznościowych. Dostęp do wielu prywatnych danych stwarza potężne możliwości, ale również zagrożenia. Wielu, o ile nie większość, użytkowników portali społecznościowych nie chce dopuścić do wykorzystania ich danych prywatnych, co więcej, nie potrafią on też zadbać o odpowiednie zabezpieczenie tych danych. Nie wszyscy mają też świadomość tego, w jaki sposób dane wrażliwe mogą być wykorzystane. Prowadząc badania mediów społecznościowych, bardzo łatwo jest dotrzeć do prywatnych danych i je niechcący upublicznić. Dlatego w tego typu badaniach warto rozważyć zastosowanie rozwiązań zmniejszających ryzyko naruszenia prywatności danych.

1. Dotychczasowy stan wiedzy

Zagadnienia zapewnienia dostępu do pewnego zbioru informacji i ograniczenia dostępu do innego zbioru znane są jako kontrola dostępu i zostały dogłębnie zbadane w obszarach baz danych [por. Bertino, Sandhu, 2005]. Stworzono wiele modeli zapewnienia bezpieczeństwa danych w bazach danych, w szczególności wielopoziomowy system zabezpieczenia baz danych [Stachour, Thuraisingham, 1990], który zaprojektowano jako środek wymuszenia kontroli dostępu i współdzielenia danych. Jednak metody te nie zawsze można zastosować w mediach społecznościowych, ich skuteczność bywa ograniczona z powodu trudności jednoznacznego ustalenia, które dane są prywatne, a które publiczne.

W badaniach mediów społecznościowych pojawia się element odkrywania wiedzy, która nie została wprost podana, ale możliwa jest do wywnioskowania na podstawie powiązanych danych. W ten sposób można odkryć prywatne dane, bazując na dostępnych danych publicznych. Metody odkrywania wiedzy zwiększają to ryzyko poprzez automatyzację operacji, co staje się również obszarem badań [por. Agrawal, Srikant, 2000]. Thuraisingham i in. [2016] opracowali koncepcję ram semantycznych i ontologii związanych z prywatnością danych w mediach społecznościowych.

Jeden obszar zagadnienia jest podobny do badań danych statystycznych, gdzie można pozwolić na dostęp do danych zagregowanych, a zabronić dostępu do danych na poziomie jednostkowym [por. Motro, Marks, Jajodia, 1994; www 1]. Wypracowano w tym obszarze badań również techniki pozwalające na kontrolę niepożądanych problemów z odkrywaniem danych. Niemniej problemem nadal pozostaje określenie stopnia agregacji oraz identyfikacja, od jakiego poziomu dane przestają być prywatne i stają się publiczne. Jednym słowem, wiadomo, jak ograniczać dostęp do danych, nie wiadomo jednak, jakie są granice bezpiecznego pozyskiwania danych. W badaniach mediów społecznościowych czasem może wystąpić odwrotny problem. W pewnych sytuacjach można zezwolić na dostęp do prywatnych danych pojedynczych osób, a zabronić agregowania tych informacji.

W badaniach w dziedzinie bezpieczeństwa internetowego prowadzone są prace, które mają na celu głównie zabezpieczenie się przed wykradzeniem prywatnych danych poprzez włamania i przejęcie kont użytkowników, odgadnięcie haseł itp. Prace te w mniejszym stopniu skupiają się na bezpieczeństwie świadomie opublikowanych informacji, które to w portalach społecznościowych są powszechnie publikowane. Niektóre osoby umieszczają na portalach społecznościowych zdjęcia swoich dowodów osobistych, kart kredytowych, praw jazdy,

ujawniając przy okazji np. PESEL, imiona swoich rodziców czy nazwisko panieńskie matki. Działania takie pozwalają na zdobycie tych cennych informacji w sposób bardzo łatwy, bez potrzeby naruszania zabezpieczeń.

2. Wybrane aspekty prywatności i bezpieczeństwa danych

Prywatność w mediach społecznościowych można rozpatrywać w wielu aspektach, lecz z punktu widzenia badań mediów społecznościowych najistotniejsze wydają się pytania:

- Jakie informacje są i mogą być gromadzone?
- Jak te informacje są współdzielone pomiędzy użytkownikami?
- Jak portale społecznościowe mogą dystrybuować dane użytkowników do innych podmiotów, w tym badaczy?
- Jakie informacje mogą być wywnioskowane metodami odkrywania wiedzy na podstawie upublicznionych danych?

Rejestrując się na portalach społecznościowych, użytkownicy zwykle udostępniają swoje podstawowe dane, takie jak e-mail, imię, nazwisko, data urodzenia, miejsce zamieszkania. Często też udostępniają dodatkowe dane, np. numer telefonu, zdjęcie, adres, miejsce pracy, swoją domową stronę internetową czy relacje z innymi osobami. Głównym celem istnienia mediów społecznościowych jest współdzielenie informacji pomiędzy ich użytkownikami. Uczestnicy chcą jednak mieć kontrolę nad tym, kto może widzieć, co oni udostępniają. Pewna część udostępnianych informacji, jak np. komentarze, recenzje, oceny, opinie, nie są danymi wrażliwymi bądź nie zawierają danych osobistych, jednak istnieje część informacji, takich jak zdjęcia, prywatne wiadomości, dane kontaktowe, które mogą odkryć więcej na temat danych osób i ich związków z innymi. Użytkownicy zwykle mogą zastrzec, kto ma dostęp do tych materiałów poprzez mechanizmy prywatności portali.

Kolejnym zagadnieniem jest, co portale społecznościowe mogą robić z pozyskanymi danymi. Nie jest niezwykłe, że mogą one agregować, dystrybuować, udostępniać, sprzedawać dane i treści użytkowników. Powiązane z wymienionymi jest zagadnienie dostępności informacji. Gdy informacja stanie się raz dostępna publicznie, może ona być przechowywana, archiwizowana na wielu różnych stronach internetowych. Może stanowić to naruszenie prywatności, gdy informacje są rozpowszechniane bez zgody właściciela. Użytkownicy zwykle mają prawo do usunięcia swoich informacji, jednak nie zawsze jest ono skuteczne w praktyce.

Informacje z portali społecznościowych stają się źródłem działań prawnych. Znane są przypadki zwolnienia z pracy po nieprzychylnych wypowiedziach w sieci czy po opublikowaniu swoich zdjęć. Również firmy ubezpieczeniowe szukają na portalach społecznościowych dowodów na nadużycia. Powołując się na [Keenan, 2011], Facebook był cytowany w 33% wniosków rozwodowych w roku 2011. Nowym trendem jest żądanie przez przyszłych pracodawców loginów i haseł do Facebooka od kandydatów do pracy, jak i również już zatrudnionych pracowników. Metody odkrywania wiedzy pozwalają na wydobycie na podstawie dostępnych danych informacji, co do których nie było zamiaru jej udostępnienia. Rozwój technologii i upowszechnienie Internetu sprawiły, że przetwarzanie danych osobowych stało się obecnie stosunkowo łatwe i szybkie. W efekcie dla wielu przedsiębiorstw wykorzystanie danych osobowych stało się podstawą ich modeli biznesowych. Narzędzia odkrywania wiedzy wykorzystują zaawansowane techniki komputerowe w celu wykrycia uprzednio nieznanych wzorców i zależności w dużych zbiorach danych. Może to stanowić poważne zagrożenie bezpieczeństwa czy prywatności informacji. Wymienione trendy wykazują, że kontrola prywatności informacji osobistych staje się coraz bardziej istotna, a ryzyka z nią związane są coraz większe.

3. Możliwości ochrony bezpieczeństwa danych w popularnych portalach społecznościowych

Media społecznościowe ułatwiają publikowanie i współdzielenie informacji. Przed gwałtownym rozwinięciem się mediów społecznościowych w celu publikacji informacji ludzie tworzyli osobiste strony internetowe. Jednak takie strony były znacznie trudniejsze do znalezienia w sieci oraz istniała duża bariera technologiczna publikacji. Media społecznościowe rozwiązały te problemy poprzez usuniecie barier publikacji oraz centralizację wyszukiwania zasobów.

Zasady publikacji i prywatności danych zmieniały się na przestrzeni czasu istnienia najpopularniejszych portali społecznościowych. Dostępność opcji zarządzania prywatnością znacznie różni się pomiędzy portalami. W artykule przedstawiono zasady prywatności trzech znaczących portali: Facebook, Twitter i LinkedIn.

3.1. Facebook

Portal Facebook został uruchomiony w 2004 r. i w pierwszym okresie był adresowany do studentów uniwersytetów. Publicznie dostępny stał się w 2006 r. W pierwszym okresie większość informacji użytkownika widoczna była tylko

dla jego znajomych. Z czasem domyślne ustawienia bezpieczeństwa zmieniały się w kierunku większego ich upublicznienia. Widoczność danych jest w głównej mierze zdeterminowana poprzez domyślne ustawienia profilu na Facebooku. Istnieje wiele ustawień prywatności, które mogą być zmienione przez użytkowników. Badania [Madden, 2012] wskazują, że 58% dorosłych i 62% młodzieży dostosowuje swoje ustawienia prywatności w portalach społecznościowych. Niestety badania te pokazują również, że około 40% użytkowników nie zmienia swoich ustawień prywatności.

Facebook był oskarżany o sprzedawanie danych na temat swoich użytkowników oraz o nieusuwanie danych po skasowaniu konta. Dane te są tylko oznaczane jako usunięte, a fizycznie przez długi czas mogą znajdować się na serwerach firmy.

3.2. Twitter

Twitter został uruchomiony w 2006 r. Jego użytkownik może wysyłać i odczytywać tzw. tweety – krótkie wiadomości tekstowe (maksimum 140 znaków), wyświetlane na profilu autora wpisu oraz pokazywane użytkownikom, którzy obserwują dany profil. Z punktu widzenia dostępu do danych Twitter ma najbardziej liberalną politykę. Domyślnie wszystkie posty – tweety są widoczne publicznie. Stanowi to duże udogodnienie dla badaczy danych, ale niesie ze sobą również znaczne zagrożenia prywatności. Również do 2012 r. zasady gromadzenia danych z Twittera były relatywnie liberalne. Aplikacja gromadzi dane osobowe użytkowników i może współdzielić się nimi z innymi podmiotami.

3.3. LinkedIn

Serwis założony w grudniu 2002 r. i uruchomiony w maju 2003 r. Jest to portal społecznościowy ukierunkowany na tworzenie sieci biznesowych i udostępnianie ofert pracy. Głównym źródłem dochodu LinkedIn jest sprzedaż danych osobowych swoich użytkowników na potrzeby działań marketingowych oraz firmom zajmującym się rekrutacją pracowników.

3.4. Porównanie zasad prywatności w mediach społecznościowych

Użytkownicy, tworząc konta na portalach społecznościowych, podają swoje prywatne dane. Najczęściej są to e-mail, imię i nazwisko. Ponadto użytkownicy często są proszeni o podanie lokalizacji (adresu lub tylko miasta czy kraju),

zdjęcia, daty urodzin i często wielu innych danych. W ustawieniach prywatności użytkownicy mogą ustalić, które dane są widoczne dla innych członków portalu. Nie mają wpływu na to, że wszystkie podane dane dostępne są dla właściciela portalu, który może zarządzać nimi według zasad ustalonych w swojej polityce prywatności. Tabela 1 przedstawia podstawowe dane i sposób ich wykorzystania przez analizowane portale społecznościowe.

Tabela 1. Sposób przetwarzania danych prywatnych w wybranych portalach społecznościowych

	Facebook	**Twitter**	**LinkedIn**
1	*2*	*3*	*4*
Jakie dane osobowe są gromadzone?			
Imię i nazwisko	X	X	X
E-mail	X	X	X
Telefon	X	X	X
Lokalizacja (bieżąca)	X	X	X
Miejsce zamieszkania	X	X	X
Poprzednie miejsca zamieszkania	X		
Zdjęcie(a)	X	X	X
Witryna internetowa		X	X
Data urodzenia	X	X	X
Kontakty (książka adresowa)	X	X	X
Rodzina i związki	X		
Wykształcenie	X		X
Ukończone szkoły	X		X
Miejsce pracy	X		X
Kwalifikacje zawodowe	X		X
Języki obce	X		X
Przekonania religijne	X		
Poglądy polityczne	X		
Wydarzenia z życia (jak zawarcie związku itp.)	X		
Powiązania z innymi użytkownikami	X		X
Informacje o płatnościach	X		X
Informacje o urządzeniach (telefony, komputery)	X	X	X
Adres IP	X	X	X
Dane z postów (tekst, zdjęcia, filmy, linki)	X	X	
Kalendarz			X
Tematy e-maili			X
Wynagrodzenie			X
Sposób zbierania informacji o użytkownikach			
Od użytkownika w procesie rejestracji	X	X	X
Od partnerów zewnętrznych	X		X
Na podstawie informacji od innych użytkowników	X	X	X
Z zachowań użytkowników (na podstawie logów itp.)	X	X	
Z aplikacji i witryn internetowych	X	X	
Sposób wykorzystania informacji			
Oznaczanie na zdjęciach	X		
Meldowanie w lokalizacjach	X	X	
Sugestie kontaktów	X		X
Oferty pracy i kontakty z rekruterami			X

cd. tabeli 1

1	*2*	*3*	*4*
Sposób udostępniana informacji			
Publicznie (według ustawień prywatności)	X	X	X
Inny użytkownicy portalu (według ustawień prywatności)	X	X	X
Aplikacje i witryny firm współpracujących	X	X	X
Firmy współpracujące	X	X	X
Do celów marketingowych	X	X	X
Wyświetlanie reklam	X	X	X
Klienci firmy			X
Do celów badawczych	X	X	X

Źródło: Opracowanie własne na podstawie: [www 2; www 3; www 4].

Polityki prywatności badanych portali mogą być, jak w przypadku [www 2], napisane w prostym, zrozumiałym języku polskim, mogą być napisane w języku angielskim, prawniczym językiem uzupełnionym jednak prostymi wyjaśnieniami [www 4] lub w najgorszym przypadku tylko tekstem prawnym, w języku angielskim, bez łatwego do odczytania wyjaśnienia [www 3].

4. Deanonimizacja danych w portalach społecznościowych

Pomimo istnienia wielu ustawień prywatności danych pozostanie anonimowym w Internecie jest nadal trudne. Polityki prywatności zwykle dotyczą takich prywatnych danych osobistych, jak imię, nazwisko, zdjęcia, numer telefonu, e-mail, data urodzenia, natomiast nie obejmują informacji niezwiązanych wprost z daną osobą, np. kod pocztowy, miasto czy szkoła lub miejsce pracy. Dane te same z siebie nie są prywatne, lecz w połączeniu z innymi mogą pozwolić na identyfikację osoby. Ponadto użytkownicy mogą chcieć podzielić się pewnymi informacjami o sobie, a ukryć inne. Powstają już mniej lub bardziej wyszukane techniki data mining pozwalające na wywnioskowanie informacji, których użytkownik nie chciał udzielić.

Deanonimizacja może być rozumiana jako proces umożliwiający odkrycie tożsamości użytkownika sieci. Możliwe jest zdobycie wystarczających do identyfikacji informacji na podstawie danych wprowadzonych przez użytkownika i ewentualnie innych dostępnych zbiorów danych. Pewne dane są zwykle łatwe do zidentyfikowania, np. płeć, stan cywilny, przedział wieku, przybliżona lokalizacja itp. Mając podane i odkryte dane, można przeglądać inne bazy danych, np. bazy klientów sklepów internetowych, w celu dokładnej identyfikacji.

Dobrym przykładem jest fakt, że anonimowe konta na Twitterze i Instagramie dyrektora FBI zostały zdeanonimizowane tylko podstawie informacji, że ma

on 9 obserwujących oraz po znajomości imion jego rodziny [www 5]. Drugim spektakularnym przykładem może być odkrycie lokalizacji tajnych baz wojskowych, a nawet identyfikacja osób tam znajdujących się na podstawie global heatmap portalu Strava [www 6]. Portal ten służy do prezentacji treningów kolarskich, biegowych i innych, zarejestrowanych przy pomocy aplikacji w urządzeniach wyposażonych w GPS. Najczęściej są to smartfony lub zegarki sportowe. Zarejestrowana trasa treningu może być publiczna, dostępna tylko dla znajomych lub całkowicie prywatna. Firma Strava na podstawie ponad miliarda tras i trzech trylionów punktów tworzy co roku światową mapę miejsc, ścieżek, którymi poruszali się trenujący. Teoretycznie mapa zawiera tylko zagregowane dane wszystkich użytkowników, więc nie powinna stanowić naruszenia prywatności. Jednak po jej opublikowaniu analitycy wojskowi zauważyli, że jest ona wystarczająco dokładna, by zidentyfikować lokalizację jednostek wojskowych i szpiegowskich. Dziennikarze norweskiego portalu NRK [Lied, 2018] przeprowadzili nawet eksperyment pozwalający na zidentyfikowanie imion i nazwisk przybywających tam osób.

Sytuacja wygląda jeszcze poważniej, gdyż istnieją narzędzia i portale internetowe zajmujące się masowym pozyskiwaniem danych na temat internautów. Przykładem tego może być oprogramowanie Maltego firmy Paterva, które służy do „open source intelligence", czyli wydobywania i korelowania publicznie dostępnych danych na temat danej osoby. Podobnie portal Locatefamily [www 7] gromadzi dane z wielu dostępnych źródeł, których firma nie ujawnia i publikuje je, pozwalając odnaleźć setki tysięcy osób. Dane te pochodzą prawdopodobnie z publicznie dostępnych rejestrów (wiele państw udostępnia tego typu dane) oraz zakupionych danych zbieranych przez firmy marketingowo-reklamowe. Przykłady odkrywania wrażliwych danych z portali społecznościowych można mnożyć. Wszystkie one wskazują, że jest to bardzo poważnie zagrożenie prywatności.

5. Możliwe rozwiązania bezpieczeństwa badanych danych

Chcąc zapewnić bezpieczeństwo i prywatność danych, przy jednoczesnej możliwości ich analizy do celów naukowych czy biznesowych, można rozważyć zastosowanie wybranych z przedstawionych metod.

5.1. Ograniczenie dostępu

Ograniczenie dostępu może polegać na zabronieniu użytkownikom pozyskiwania dużych zbiorów danych. Można również zezwolić na dostęp do losowo lub według innej strategii wybranego podzbioru dużego zbioru danych. Obniży to wiarygodność uzyskanych wyników, gdyż nie będą one reprezentowały całego zbioru, lecz tylko wybrany podzbiór danych. Taką politykę stosuje Twitter do dostępu do danych. W swoim API Twitter [www 8] wprowadził ograniczenia prędkości pobierania danych oraz liczby pobieranych danych, udostępniając tylko część danych spełniających kryteria zapytania.

5.2. Rozmywanie danych

Dane mogą być (lekko) modyfikowane. Zmieniając nieco wartości pojedynczych transakcji, można zapobiec wykorzystaniu ich do odkrywania wiedzy, nadal zachowując jednak ich użyteczność do celów, dla których zostały opublikowane. Tę strategię stosuje np. U.S. Census Bureau [www 9] w publikowanych przez siebie danych.

5.3. Eliminacja niepożądanego grupowania danych

Dane często zawierają niepożądane informacje pozwalające je pogrupować. Na przykład numery rejestracyjne pozwalają pogrupować dane według miast i powiatów. Podobnie numery PESEL zawierają informacje o dacie urodzenia. Wewnętrzne numery telefoniczne w firmach i urzędach są zwykle przydzielane według rozmieszczenia pomieszczeń. Może to mieć implikacje związane z bezpieczeństwem informacji i być wykorzystane do pozyskania dodatkowej wiedzy. Dla przykładu, zbliżone wewnętrzne numery telefonów wskazują osoby pracujące w tych samych działach czy projektach. Rozwiązaniem może być przydzielanie losowych numerów tam, gdzie jest to oczywiście możliwe.

5.4. Sztuczne powiększanie zbiorów danych

W wielu przypadkach można dodać coś do danych, nie zmniejszając ich użyteczności. Przy normalnym, zalecanym sposobie korzystania z danych dodane dane nie mają żadnego znaczenia. Dodatkowe dane zwiodą tylko wykorzystujących dane w nieprawidłowy sposób. Na przykład książka telefoniczna może

być powiększona o dodatkowe fikcyjne osoby i numery telefonów. Szukając numeru telefonicznego wybranej osoby lub działu, zwróci ona poprawną odpowiedź, ale zapytania przeszukujące wszystkie osoby w dziale zwrócą dodatkowe, nieprawidłowe odpowiedzi.

5.5. Audyt zapytań o dane

Podczas gdy poprzednie metody dotyczyły publikacji danych, audyt dotyczy analizy sposobu wykorzystywania danych. Audyt może polegać na sprawdzeniu, kto i w jaki sposób korzystał z danych. Ma to działanie odstraszające, zapobiegające nadużyciom. Dzięki audytowi można wykryć nieprawidłowe korzystanie z danych, do których pozyskujący ma uprawnienia. Takim przykładem było masowe pozyskiwanie danych osobowych z bazy PESEL przez kancelarie komornicze. Komornicy mogą sprawdzać te dane, bo pozwala im na to prawo. Wykorzystują uzyskane informacje, prowadząc postępowania, np. sprawdzając aktualny adres dłużnika. Wiele kancelarii pozyskiwało na masową skalę dane niezwiązane z prowadzonymi przez nie sprawami [Kolińska-Dąbrowska, 2016].

Przedstawione rozwiązania stanowią tylko wierzchołek góry lodowej problemów ochrony prywatności danych osobowych. Zapobieganie niepożądanemu odkrywaniu wiedzy z danych wymaga wielu dalszych prac.

Podsumowanie

Dostępność dużej ilości prywatnych danych w mediach społecznościowych stwarza znaczne możliwości ich analizy. W niewłaściwych rękach może stanowić to istotne zagrożenie prywatności. Rozwiązaniem nie jest ograniczanie dostępu do danych, media społecznościowe powstały w celu wymiany takiej informacji. Należy poszukiwać pośrednich metod, co wskazano w artykule jako możliwe rozwiązania.

Na kierunki dalszych prac na pewno wpływ będzie miało Rozporządzenie o Ochronie Danych Osobowych (RODO) [www 10], które weszło w życie 25 maja 2018 r. Po jego wprowadzeniu każdy podmiot przetwarzający dane zobowiązany będzie bezpośrednio stosować zawarte tam przepisy. Mimo iż dotyczą one głównie ochrony przed wyciekiem danych, niezbędna okaże się jednak ocena wpływu przetwarzania danych na ochronę praw i wolności osób, których one dotyczą.

Literatura

Agrawal R., Srikant R. (2000), *Privacy-preserving Data Mining*, "Sigmod", Vol. 29, Iss. 2, s. 439-450.

Bertino E., Sandhu R. (2005), *Database Security-concepts, Approaches, and Challenges*, "IEEE Transactions on Dependable and Secure Computing", Vol. 2(1), s. 2-19.

Keenan M. (2011), *Alarming Increase in Facebook Related Divorces in 2011*, https://www.divorce-online.co.uk/blog/alarming-increase-in-facebook-related-divorces-in-2011/ (dostęp: 26.06.2018).

Kolińska-Dąbrowska M. (2016), *Wielkie zasysanie danych PESEL przez komorników*, http://wyborcza.pl/1,155287,20605317,wielkie-zasysanie-danych-pesel-przez-komornikow.html (dostęp: 26.06.2018).

Lied H. (2018), *How We Found the Identity of Military Personnel Using Strava*, https://nrkbeta.no/2018/01/31/how-we-found-the-identity-of-military-personnel-using-strava/ (dostęp: 26.062018).

Madden M. (2012), *Privacy Management on Social Media Sites* [w:] *Pew Internet Report*, Pew Research Center, Washington, DC, USA, s. 1-20.

Motro A., Marks D.G., Jajodia S. (1994), *Aggregation in Relational Databases: Controlled Disclosure of Sensitive Information* [w:] *European Symposium on Research in Computer Security*, Springer, Berlin–Heidelberg, s. 429-445.

Stachour P.D., Thuraisingham B. (1990), *Design of LDV: A Multilevel Secure Relational Database Management System*, "IEEE Transactions on Knowledge and Data Engineering", Vol. 2(2), s. 190-209.

Thuraisingham B., Abrol S., Heatherly R., Kantarcioglu M., Khadilkar V., Khan L. (2016), *Analyzing and Securing Social Media*, CRC Press, Boca Raton.

[www 1] Report on Statistical Disclosure Limitation Methodology (2005), Federal Committee on Statistical Methodology, https://fcsm.sites.usa.gov/reports/policy-wp/ (dostęp: 31.01.2018).

[www 2] Zasady dotyczące danych, Facebook, https://www.facebook.com/privacy/explanation (dostęp: 05.02.2018).

[www 3] Twitter Privacy Policy, Twitter, https://twitter.com/en/privacy (dostęp: 05.02.2018).

[www 4] Privacy Policy, LinkedIn, https://www.linkedin.com/legal/privacy-policy (dostęp: 05.02.2018).

[www 5] *Jak dyrektor FBI został zdeanonimizowany przez dziennikarkę? Kilkoma prostymi trickami*…, Niebezpiecznik, https://niebezpiecznik.pl/post/jak-dyrektor-fbi-zostal-zdeanonimizowany-przez-dziennikarke-kilkoma-prostymi-trickami/ (dostęp: 26.06.2018).

[www 6] https://labs.strava.com/heatmap/ (dostęp: 26.06.2018).

[www 7] https://www.locatefamily.com (dostęp: 26.06.2018).

[www 8] Rate Limiting, Twitter, https://developer.twitter.com/en/docs/basics/rate-limiting (dostęp: 26.06.2018).

[www 9] United States Census Bureau, https://www.census.gov (dostęp: 26.06.2018).

[www 10] Ustawa z dnia 10 maja 2018 r. o ochronie danych osobowych, http://prawo.sejm.gov.pl/isap.nsf/download.xsp/WDU20180001000/T/D20181000L.pdf (dostęp: 30.08.2018).

DATA SECURITY AND PRIVACY IN SOCIAL MEDIA RESEARCH

Summary: Social media analysis allows to reveal information in Web portals, as well as by usage of information seeking methods. That data publication is an invasion of privacy. The paper presents different problems connected with privacy invasion. Next, author focuses on conceptualization of solutions for security and privacy protection in social media research.

Keywords: social media, knowledge discovery, personal data protection.