

Andrzej Kobyliński

Szkoła Główna Handlowa w Warszawie
Instytut Informatyki i Gospodarki Cyfrowej
Kolegium Analiz Ekonomicznych SGH
e-mail: andrzej.kobyliński@sgh.waw.pl

Agregatory wiadomości – remedium na przeładowanie informacyjne?

Kody JEL: C88, D83, L86

Słowa kluczowe: agregator wiadomości, agregator treści, przeładowanie informacyjne

Streszczenie. Internet, stanowiący olbrzymi zasób informacji, powiększa się bezustannie. Prowadzi to do przeładowania informacyjnego, gdy ludzie nie są w stanie zapanować nad nadmiarem dostępnych treści. Pewnym rozwiązaniem problemu są agregatory wiadomości – narzędzia, które filtrują wiadomości pochodzące z wielu źródeł internetowych i prezentują wybrane informacje w jednym miejscu. Proces selekcji dokonywany jest przez dziennikarzy wspieranych przez algorytmy uczenia maszynowego, co prowadzi do personalizacji treści prezentowanych poszczególnym odbiorcom.

Wprowadzenie

Ilość informacji, jaką ludzie wytworzyli (i zapisali) w czasie całej historii ludzkiej cywilizacji, jest trudna zarówno do opisanego, jak i wyobrażenia. Rzetelne pod względem naukowym badanie przeprowadzone w School of Information Management and Systems w University of California (Berkeley) przez Lymana i Variana z zespołem (2000) wykazało, że w roku prowadzenia badań (2000 r.) ludzkość wytwarzała rocznie ok. 1–2 mld gigabajtów informacji (1–2 exabajta (EB), czyli 10^{18} B). Warto zauważyć, że tylko 0,003% tych informacji została wydrukowana. Dla porównania – autorzy opracowania

szacują, że 5 EB odpowiada wszystkim słowom wypowiedzianym kiedykolwiek przez ludzi¹.

Od czasu przeprowadzenia tego badania minęło 17 lat, a jest to cała era w rozwoju technologii służących do zapisu informacji. Techniczna łatwość obsługi i dostępność cenowa urządzeń, które pozwalają wytwarzać i gromadzić dane różnego rodzaju przyczyniły się do tego, że następuje bezustanny wzrost liczby zapisywanych danych². Większość utwalonych różnymi metodami danych (szczególnie zarejestrowanych w formie obrazów, filmów) jest następnie kasowana, ale pewna część wyprodukowanych danych (nie tylko zdjęć, ale również tekstów) jest jednak publikowana w internecie. Szacunki dotyczące tego, ile spośród tych informacji jest opublikowanych, są mocno rozbieżne.

Cytowane wyżej dane dotyczące wytwarzania danych, pochodzą sprzed kilkunastu lat. Obecne (styczeń 2017 r.), prowadzone na bieżąco estymacje wskazują, że indeksowanych jest przez główne wyszukiwarki ok. 5,1 mld stron (domen) internetowych (Kunder, 2017)³, chociaż tylko ok. 1,1 mld stanowią strony aktywne (Internet Live Stats, 2017). Pozostałe strony (ok. 75% całości) pozostają nieaktywne (są to głównie strony o zarejestrowanej nazwie, na których nie jest odnotowywany żaden ruch i o standardowej, szablonowej treści występującej w innych stronach przypisanych do tej samej domeny) (Netcraft, 2017).

Te olbrzymie liczby są trudne do wyobrażenia, ale lepszą optykę tych wielkości można otrzymać, gdy spróbuje się przełożyć je na wielkości fizyczne. Dewey (2015), dziennikarka Washington Post, przeprowadziła badanie kilkuset losowo wybranych stron internetowych, próbując określić, ile stron papieru formatu A4 wymagane jest do wydrukowania jednej strony internetowej. Okazało się, że potrzeba do tego przeciętnie 6,5 kartki⁴. Proste obliczenia prowadzą do wniosku, że nawet gdy przyjmiemy najmniejszą z wymienionych w poprzednim akapicie wielkości, to: $1,1 \text{ mld} \cdot 6,5 \text{ kartki} = 7 \text{ mld}$ stron wydruku, czyli gdyby wydruk ustawić na półce, to musiały ona mieć 700 km długości!

¹ Dla porównania warto wspomnieć, że 26 mln książek zgromadzonych w największej na świecie Bibliotece Kongresu Amerykańskiego odpowiada tekstowi o objętości 10 TB (10^{12} bajta) (Lyman, Varian, 2000), czyli jest 100 tys. razy mniejsza.

² Na przykład zwróćmy uwagę na to, jak wiele fotografii i filmów wykonuje się obecnie w porównaniu do stanu sprzed kilkadziesiąt, a nawet kilkunastu lat. Gdy nośniki zapisujące obrazy wymagały obróbki fotochemicznej, robienie zdjęć (nie mówiąc już o filmach na taśmie 8 mm) było nieporównywalnie rzadsze w stosunku do czasów współczesnych, gdy bez ustanku jesteśmy fotografowani i nagrywani nie tylko przez fotoamatorów, ale również przez kamery telewizji przemysłowej (CCTV).

³ Na tej samej stronie Kunder (2017) pokazuje na wykresie odmiennie dane – wynika z nich, że liczba stron internetowych przekracza 45 mld. Ponieważ nie wyjaśnia on tej rozbieżności, można przypuszczać, że 5,1 mld to liczba zarejestrowanych domen, a 45 mld to liczba subdomen.

⁴ Eksperyment przeprowadzony przez autora na kilkadziesiąt przypadkowych stronach potwierdził zgodność z tą średnią.

Inną wartą zasygnalizowania kwestią jest istnienie sieci głębokiej (ukrytej – *Deep Web*), obejmującej strony dynamicznie generowane, pobierające dane z baz danych, których nie da się znaleźć przez tradycyjne wyszukiwarki, zawartość sieci intranetowych, stron zabezpieczonych hasłem dostępu lub do dostępu wymagających specjalistycznego oprogramowania (np. TOR). Oczywiście, o wielkości tej sieci można tylko spekulować, gdyż z założenia nie mamy na ten temat wiedzy, ale uważa się, że indeksowana część zawartości sieci stanowi tylko 1% (Pagliery, 2014) do 4% (Deep Web Sites, 2017) całości zgromadzonych zasobów.

Niezależnie od olbrzymiej wielkości treści, jakie znaleźć można na stronach WWW, nie sposób nie odnieść się do innego, dynamicznego parametru, jakim można scharakteryzować internet, a mianowicie możliwości (potencjału) przesyłania informacji. Przesył danych, jaki dokonuje się w sieci w końcu 2016 roku szacowany był na 1,1 zetabajtów (10^{21} B) rocznie i oczekuje się, że w kolejnych kilku latach wzrost będzie na poziomie 22% rocznie (Cisco, 2016). W 2015 roku 70% z tych przesyłanych danych miał formę strumienia wideo, szacunki mówią, że w ciągu najbliższych 5 lat nastąpi w tym zakresie wzrost do 82% (Cisco, 2016).

Ciekawe dane publikowane są na bieżąco w portalu Internet Live Stats (2017), który prezentuje aktywności, jakie dokonują się w sieci w ciągu jednej sekundy – co sekundę przesyłanych jest: 41 321 GB danych, wysyłanych jest 2 557 495 e-maili (z czego ok. 67% to spam), poprzez serwis społecznościowy Tweeter wysyłane są 7482 tweety, w Instagramie udostępniane są 762 nowe zdjęcia, na platformę mikroblogową Tumblr wysyłanych jest 1201 postów, poprzez Skype inicjowanych jest 2430 rozmów, Google otrzymuje 58 330 zapytań, rozpoczyna się 67 848 sesji wideo na YouTube.

1. Przeładowanie informacyjne

Pojęcie „przeładowanie informacyjne” (*information overload*) zastosował po raz pierwszy Alvin Toffler w 1970 roku (za: Understanding Information Overload). Przewidział on, że gwałtownie rosnąca liczba wytwarzanych informacji może w przyszłości stać się problemem dla cywilizacji ludzkiej. Możliwość przetwarzania informacji przez mózg ludzki nie zmieniła się istotnie w minionych dziesiątkach lat, a w tym samym czasie liczba docierających do nas informacji uległa zwielokrotnieniu. Może to prowadzić do stresu, gdy człowiek ma poczucie nadmiaru wpływających informacji, podczas gdy nie ma zdolności, umiejętności ani sprawdzonych metod, by informacje te wykorzystać w procesie podejmowania decyzji i finalnie – w działaniu.

Przyczyny przeładowania informacyjnego są różnorodne. Część z nich ma charakter obiektywny: łatwy dostęp do internetu z mnóstwem łatwo dostępnych i niekiedy atrakcyjnie podanych informacji, łatwe w użyciu metody komunikowania się ze znajomymi, czy to przez wysyłanie maili, czy poprzez przesyłanie postów poprzez media społecznościowe, ale również praktycznie bezkosztowy sposób duplikowania danych.

Powoduje to, że ludzie często rozsyłają informacje do wszystkich potencjalnie zainteresowanych, a nie, jak to było dotychczas, wyłącznie do tych, dla których informacja taka jest z pewnością istotna.

Do powodów subiektywnych można zaliczyć słabą jakość informacji dostępnych bezpłatnie online, które są często nadmiernie rozbudowane (przegadane), napisane niepoprawnym, niechlujnym, utrudniającym zrozumienie językiem, często zawierające błędy faktograficzne lub niespójności, które wymagają dalszych poszukiwań w celu ich potwierdzenia bądź zakwestionowania. Nie zawsze zła jakość informacji spowodowana jest przez przypadek bądź nieuwagę. Niekiedy informacje fałszywe produkowane są z premedytacją, w celu przysporzenia korzyści ich autorom. W licznych przypadkach jakość informacji jest celowo obniżana, gdyż są one wtedy łatwiejsza do sprzedaży i przynosi ich dostawcom większe zyski (Fazlagić, 2010). Z drugiej strony w mediach masowych informacje poważne (polityka, gospodarka) wypierane są przez trzeciorzędne pseudonewsy (katastrofy, przestępstwa, celebryci, „wyciskacze łez”) – informacja ulega tabloidyzacji. Niektórzy socjologowie przekonują, że żyjemy w epoce ignorancji, a z całą pewnością przyczynia się do tego internet, w którym trudno o hierarchię informacji i źródeł (Keen, 2007).

2. Sposób korzystania z informacji w sieci

Problem nadmiaru informacji wartej przyswojenia (ew. braku czasu na przeczytanie wszystkich interesujących źródeł lub brak pieniędzy na zakup wszystkich ciekawych materiałów) nie jest nowym zjawiskiem. W czasach sprzed internetu, czyli wyłącznie periodyków drukowanych, problem starały się rozwiązywać czasopisma, które przedrukowywały materiały z innych, oryginalnych źródeł. Pierwszym na świecie tego rodzaju czasopismem był amerykański miesięcznik „Reader’s Digest” (Reader’s Digest, 2016), wydawany obecnie w 49 edycjach narodowych i 21 językach, o ogólnym nakładzie wynoszącym 10,5 mln egzemplarzy. Na rynku polskim czasopismem tego rodzaju, które ukazuje się od 1965 roku, jest dwutygodnik (poprzednio tygodnik) „Forum”; obecnie najpopularniejszym czasopismem mającym formułę przeglądu prasy krajowej i światowej jest „Tygodnik Angora” o niezwykłej popularności – plasującym się na 6. miejscu pod względem sprzedaży tygodników w Polsce (270 tys. w listopadzie 2016 r. wg (ZKDP, 2016)) (wyprzedzają go wyłącznie magazyny telewizyjne i tzw. żółta prasa) a wyprzedza on wszystkie tzw. tygodniki opinii, z których czerpie większość poważnych materiałów.

Korzystanie z informacji w sieci przez przeciętnego internautę było (i w większości przypadków nadal jest) wyjątkowo proste. Po włączeniu przeglądarki należy wejść na stosowną stronę www, wpisując właściwy adres lub wybierając adres ulubionej strony z zapamiętanych zakładek. Ale jednocześnie, mimo że przeciętny internauta w Polsce spędza na samym surfowaniu od ok. 2 godz. (CBOS, 2015) do 2,5 godz. dziennie (Gemius, 2015), większość z nas ma poczucie, że każdego dnia omijają nas ważne wie-

ści, lub że tracimy zbyt wiele czasu na żmudne przeglądanie portali internetowych w poszukiwaniu nowinek na różne tematy. Dlatego ten tradycyjny, oczywisty sposób docierania do ulubionych stron stopniowo uzupełniany jest o inne metody.

3. Agregatory wiadomości

Coraz więcej osób, zamiast samodzielnego poszukiwania i odkrywania informacji, wykorzystuje do tego celu narzędzia wspomagające. Wprawdzie ich użyteczność może być kwestionowana, niemniej jest ich coraz więcej i zdobywają coraz szerszą popularność. Narzędzia takie noszą ogólną nazwę agregatorów wiadomości (również: treści, newsów – *news aggregators*). Ich zasada działania jest taka sama, jak w przypadku czasopism mających charakter przeglądów prasy: agregują one co ciekawsze wiadomości pochodzące z wielu źródeł internetowych (prasy online, blogów, wideoblogów, podcastów audio i wideo) i prezentują wybrane informacje na jednej stronie w uporządkowanej, ujednocionej formie, umożliwiającą ich łatwe przeglądanie. Forma ta dla każdej wybranej wiadomości ogranicza się zwykle do zdjęcia obrazującego temat, nośnego tytułu, zdania lub akapitu sygnalizującego zawartość treściową, link do źródła wiadomości i ewentualnie ocenę czytelników.

Podobnie jak to jest w przypadku wielu innych trendów, jakie pojawiły się w ostatnich latach w obszarze IT, takich jak: ruch wolnego oprogramowania *versus* oprogramowanie własnościowe; zwinne metody wytwarzania oprogramowania *versus* metody kaskadowe; finansowanie społecznościowe *versus* tradycyjne finansowanie kapitałowe, również w odniesieniu do agregatorów wiadomości można mówić o agregacji społecznościowej i firmowej (ale jednocześnie coraz częściej spersonalizowanej).

Agregacja społecznościowa polega na tym, że na platformie agregującej pojawiają się takie wiadomości, które budzą zainteresowanie czytelników. Każdy zarejestrowany użytkownik takiej platformy może poddać pod osąd publiczny (w formie głosowania) znalezioną w internecie stronę z informacją potencjalnie interesującą dla szerszego grona czytelników. Strony/filmy/fotografie, na które padło wiele głosów pojawiają się na stronie głównej, zyskując przez to zwielokrotnioną popularność i możliwość dotarcia do mniej aktywnych obserwatorów takiej platformy – tych którzy, przykładowo, mają zwyczaj przeglądania wyłącznie newsów ze strony głównej.

Przykładami takich agregatorów o skali światowej są, na przykład Digg.com oraz Reddit.com. Popularność pierwszej z wymienionych platform, wyjątkowo duża w poprzedniej dekadzie, w ostatnich latach mocno ucierpiała i aktualnie portal plasowany jest na ok. 2000 pozycji w światowym rankingu popularności (wg rankingu Alexa ze stycznia 2017 r.), ze spadkiem o 500 miejsc w ostatnim kwartale. Reddit.com z kolei zyskuje na popularności, według Alexa jest obecnie na 23. miejscu w świecie, z tendencją rosnącą. Odpowiednikiem tych portali w Polsce jest Wykop.pl, bardzo popularny w pewnych środowiskach, według Alexa klasyfikowany na 17. miejscu w Polsce (i ok. 1400 miejscu na świecie, ze wzrostem o 500 pozycji w ostatnim kwartale). A zatem

w przypadku agregacji społecznościowej, filtrowaniem wiadomości zajmują się gromadnie sami użytkownicy.

Z kolei **agregacja firmowa** polega na tym, że firma prowadząca portal agregujący wiadomości, na stronie głównej umieszcza newsy na podstawie wiedzy, intuicji, przekonania pracowników firmy. Zwykle właściciele takiej aplikacji zastrzegają, że wyboru dokonuje „zespół doświadczonych dziennikarzy” przy wsparciu złożonych algorytmów uczenia maszynowego, na podstawie sygnałów z sieci społecznościowych. Tak ogólne stwierdzenia nie pozwalają na ocenę wewnętrznego algorytmu agregatorów, gdyż te nigdy nie są publikowane pozostając własnością firmy, można jedynie próbować ocenić trafność publikowanych rezultatów, traktując sam algorytm jako „czarną skrzynkę” o nieznanym strukturze wewnętrznej i mechanizmie dobierania newsów do publikacji. W miarę korzystania z takiego agregatora, aplikacja może „uczyć się” preferencji użytkownika, a inteligentny algorytm filtrowania treści wykorzystuje takie dodatkowe dane, jak historia przeglądanych wiadomości, aktualna lokalizacja użytkownika, dodatkowe informacje o odbiorcy pobrane z kanałów społecznościowych, jeśli ten zalogował się w agregatorze za pomocą konta w medium społecznościowym. Dalszego zindywidualizowania dostarczanych treści można dokonać manualnie, doprecyzowując obszar swoich zainteresowań, wskazać portale, z których mają być pobierane treści, jak również pozwolić aplikacji poznać lepiej nasze preferencje, oceniając przeczytane artykuły. W efekcie, dzięki sztucznej inteligencji na głównym ekranie znajdują się spersonalizowane treści, dobrane pod konkretnego użytkownika.

Rynek agregatorów tego typu znajduje się w ciągłym rozwoju. Największe firmy z branży standardowo oferują tego rodzaju aplikacje. MS Microsoft udostępnia MSN News, Google – Google News, Samsung – Upday, Apple – Apple News. Na portalach zajmujących się katalogowaniem i udostępnianiem użytkownikom aplikacji komputerowych, znaleźć można ponad 20 bezpłatnych, dostępnych w Polsce aplikacji, zaliczanych do tej grupy oprogramowania.

4. Zalety i wady agregatorów

Zalety – konsumpcja treści dostarczanych przez agregatory ma wiele zalet. Nie wymaga od użytkownika żadnych niezbędnych do tej pory działań – nie trzeba systematycznie wchodzić na swe ulubione strony, nie trzeba katalogować zakładek. Otrzymujemy gotowy materiał, początkowo może niezbyt dopasowany do naszych oczekiwań, ale stopniowo coraz trafniejszy, w miarę gromadzenia przez agregator coraz większej liczby danych na temat naszych zainteresowań. Jednocześnie, dzięki temu, że agregator wyświetla treści z wielu portali, odbiorca może mieć poczucie, że nie ograniczamy się do informacji z jednego źródła, lecz pozostajemy dobrze poinformowani, czerpiąc z wielu heterogenicznych zasobów sieciowych. Agregatory dostarczają wartościowych treści, a użytkownikom pozostaje ich konsumpcja. Jest to ogromna zaleta, gdyż odpada

nam dotychczasowa potrzeba filtrowania informacji – dostajemy wyłącznie informacje wartościowe, leżące w kręgu naszych zainteresowań.

Wady – wiara w to, że agregatory integrują informacje ze wszystkich źródeł internetowych jest bezpodstawna. Integratory firmowe udostępniają wyłącznie publikacje z portali, z którymi zawarta została stosowna umowa licencyjna, tak więc większość witryn nie jest uwzględniana w wynikach integracji. Powoduje to oczywiste zubożenie wartości udostępnianych wyników. W efekcie zamykamy się w zbiorze kilkunastu serwisów, z których filtrowane są dostarczane nam treści. Cały przeogromny internet redukujemy do jego mikroskopijnej namiastki. Ograniczenie to nie dotyczy agregatorów społecznościowych, lecz tam z kolei aktywnie działa i podbija wyniki grupa osób o specyficznych zainteresowaniach, dlatego na stronach głównych tych portali dominują sensacje, śmieszne filmiki i ciekawostki.

Również możliwość parametryzacji agregatorów, polegająca na możliwości wskazania źródeł, z których mają być czerpane wiadomości, jest nader często niewykorzystywana. Z obserwacji autora wynika, że znaczna część użytkowników tych narzędzi nie zdaje sobie sprawy z tego, że taka możliwość istnieje, a jeśli nawet wiedzą, to z lenistwa nie wskazują ulubionych źródeł, zdając się na gust pracowników portalu.

Taka przyjmowana bezkrytycznie postawa zaufania wobec dostawców treści jest nieuzasadniona. Autor przez dwa tygodnie prowadził badanie polegające na tym, że codziennie rejestrował wszystkie najważniejsze wiadomości dnia, jakie ukazywały się na stronach głównych dwóch popularnych agregatorów treści. Agregatory nie zostały spersonifikowane, więc na stronie głównej ukazywały się newsy autoryzowane przez wydawców portali. Notowane były źródła pojawiających się wiadomości oraz rodzaju wiadomości (z podziałem na dwie kategorie: wiadomość „polityczna” i „neutralna politycznie”). Źródła wiadomości zostały przez autora arbitralnie zaliczone do dwóch kategorii: „liberalnych” i „konserwatywnych”. W obu portalach formalnie (ilościowo) zachowana została neutralność polityczna: połowa pokazywanych wiadomości pochodziła ze źródeł „liberalnych”, druga połowa – „konserwatywnych”. Analiza pod kątem rodzaju wiadomości dała jednak wyniki bardzo znaczące – większość wiadomości o charakterze politycznym pochodziła ze źródeł „liberalnych”, a wiadomości politycznie neutralne (np. wypadki, katastrofy, wydarzenia z odległych części świata) publikowane były zwykle za źródłami „konserwatywnymi”. Próba spersonalizowania przez dodanie dodatkowych źródeł wiadomości też nie mogła dać zadawalających wyników – na liście tygodników (lub ich internetowych mutacji), które zasilają system, nie ma możliwości wyboru prawie żadnych tygodników opinii, niezależnie od ich opcji (co prawdopodobnie wynika z sygnalizowanego braku licencji na korzystanie ze źródła). Jednostronnie wygląda możliwość dodania źródła wiadomości regionalnych – na liście do wyboru jest 7 lokalnych gazet regionalnych i aż 23 mutacje jednego dziennika ogólnopolskiego.

Warto sobie uświadomić, że polegając na mechanizmach agregującym informacje, mamy do czynienia nie z pojedynczym, lecz podwójnym filtrem spływających do czytelnika treści. Pierwszy filtr występuje zawsze – niezależnie od tego czy korzystamy

z agregatora, czy samodzielnie przeszukujemy internet. Stanowi go osoba, która publikuje wiadomość w sieci na witrynie (autor bloga, dziennikarz internetowy, redaktor odpowiedzialny, wydawca) – decydując nie tylko o samej jego treści, ale umiejscowieniu na stronie, chwytliwym tytule, atrakcyjnym zdjęciu i wciągającym leadzie. Drugim filtrem, tym razem występującym już tylko w agregatorach, jest opisany wyżej mechanizm agregacji. Czytelnik ostatecznie uzyskuje już podwójnie przefiltrowaną treść.

Podsumowanie

Internet to olbrzymi zasób aktualnych informacji. Źródeł interesujących tematów jest wiele. Świadomy użytkownik lubi wiedzieć dużo, być na bieżąco z interesującymi go wątkami. Zaczyna być przeładowywany informacjami, treści jest tak wiele, że nie są możliwe do przyswojenia. I tu pojawia się narzędzie, ułatwiające życie człowiekowi – aplikacja agregatora treści, która wyławia z sieci informacje, które mogą być interesujące dla odbiorcy i prezentuje mu je w skondensowany sposób. Co więcej – dzięki zastosowaniu sztucznej inteligencji (i dodatkowo bezpośrednim wskazaniom użytkownika) uczy się preferencji odbiorcy i bezustannie doskonali trafność swoich wskazań. Ale pomimo niewątpliwych zalet, nie można zapominać o niedoskonałościach takich narzędzi. Agregatory pobierają informacje z ograniczonej liczby źródeł, a ich wskazania mogą być manipulowane przez wydawcę. Korzystając z nich uzyskujemy wygodę, ale jednocześnie godzimy się na narzucenie sobie pewnej formy cenzury. Sami, dla własnej wygody, ograniczamy przed sobą horyzonty.

Bibliografia

- Apple News. Pobrano z: <http://www.apple.com/news/> (23.01.2017).
- CBOS (2015). *Internauci*, 90. Pobrano z: http://www.cbos.pl/SPISKOM.POL/2015/K_090_15.PDF (21.01.2017).
- Cisco (2016). *The Zettabyte Era: Trends and Analysis (white paper)* (2.06.2016). Pobrano z: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.pdf> (19.01.2017).
- Deep Web Sites. Pobrano z: <https://www.deepweb-sites.com/how-big-is-the-deep-web/>.
- Dewey, C. (2015). If you could print out the whole Internet, how many pages would it be? *The Washington Post*, 18.05.2015. Pobrano z: https://www.washingtonpost.com/news/the-intersect/wp/2015/05/18/if-you-could-print-out-the-whole-internet-how-many-pages-would-it-be/?utm_term=.e5945232822f (19.01.2017).
- Digg, digg.com (23.01.2017).
- Fazlagić, J.A. (2010). Zjawisko „nadmiaru informacji” a współczesna edukacja. *E-mentor*, 4 (36). Forum. Pobrano z: <http://www.forumdwutygodnik.pl/> (22.01.2017).

- Gemius (2015). *Polski internauta w sieci spędza średnio 36 dni rocznie* (17.07.2015). Pobrano z: <https://www.gemius.pl/wszystkie-artykuly-aktualnosci/polski-internauta-w-sieci-spe-dza-srednio-36-dni-rocznie.html> (21.01.2017).
- Google News. Pobrano z: <https://news.google.pl/> (23.01.2017).
- Internet Live Stats. Pobrano z: <http://www.internetlivestats.com/>.
- Keen, A. (2007). *Kult amatora. Jak internet niszczy kulturę*. Warszawa: Wydawnictwa Akademickie i Profesjonalne.
- Kunder M. de. *The size of the World Wide Web (The Internet)*. Pobrano z: <http://www.worldwidewebsite.com/> (19.01.2017).
- Lyman, P., Varian, H.R. (2000). *How Much Information?*. Berkley: University of California
Pobrano z: <http://www2.sims.berkeley.edu/research/projects/how-much-info/how-much-info.pdf> (18.01.2017).
- MSN News. Pobrano z: <https://www.msn.com/en-us/news> (23.01.2017).
- Netcraft. *How many active sites are there?* Pobrano z: <https://www.netcraft.com/active-sites/> (19.01.2017).
- Pagliery, J. (2014). *The Deep Web you don't know about* (10.03.2014). Pobrano z: <http://money.cnn.com/2014/03/10/technology/deep-web/> (19.01.2017).
- Pappas, S. (2016). *How Big Is the Internet, Really?* (18.03.2016). Pobrano z: <http://www.livescience.com/54094-how-big-is-the-internet.html> (19.01.2017).
- Reader's Digest. Pobrano z: <http://www.rd.com/> (22.01.2017).
- Reddit. Pobrano z: <https://www.reddit.com/> (23.01.2017).
- Tygodnik Angora. Pobrano z: <https://www.angora.com.pl/> (22.01.2017).
- Understanding Information Overload. Pobrano z: <http://www.infogineering.net/under-standing-information-overload.htm>.
- Upday. Pobrano z: <http://upday.com/pl/> (23.01.2017).
- Związek Kontroli Dystrybucji Prasy. Pobrano z: <https://www.zkdp.pl/>, <https://www.teleskop.org.pl/zkdp/index.jsp?p=publicDataNoReg> (22.01.2017).

NEWS AGGREGATORS – A REMEDY FOR THE INFORMATION OVERLOAD?

Keywords: news aggregator, information overloading

Summary. Internet, which is a tremendous resource of information is growing constantly. This leads to information overload, when people are not able to control the excess of available content. One of the solutions to the problem are news aggregators – tools that filter news from multiple web sources and present selected information in one place. The selection process is carried out by journalists supported by the machine learning algorithms, which leads to personalize content presented to the particular customers.

Translated by Andrzej Kobyliński

Cytowanie

Kobyliński, A. (2017). Agregatory wiadomości – remedium na przeladowanie informacyjne? *Ekonomiczne Problemy Usług, 1* (126/2), 83–92. DOI: 10.18276/epu.2017.126/2-09.