

JAROSŁAW BORUSZEWSKI

Operatory logiczne w zapytaniach wyszukiwawczych

Wprowadzenie

Wyszukiwarki internetowe oferują możliwość korzystania z operatorów logicznych¹, a umiejętność ich stosowania jest uważana za istotny element kompetencji informacyjnych współczesności. *Katalog kompetencji medialnych, informacyjnych i cyfrowych* z roku 2014 podaje, że umiejętność używania operatorów logicznych w formułowaniu zapytań wyszukiwawczych należy do standardu edukacji informacyjnej. Standard ten „tworzy zespół umiejętności aktywnego i obeznanego w świecie cyfrowym obywatela” [*Katalog...*, 2014, s. 9]. W odniesieniu do operatorów logicznych brzmi on następująco: „Umie skorzystać z fraz i operatorów logicznych podczas wyszukiwania w elektronicznych źródłach informacji” [*Katalog...*, 2014, s. 22]. Stosowanie operatorów logicznych w wyszukiwaniu informacji odgrywa zatem dość znaczącą rolę w charakterystyce kompetencji informacyjnych. Nie określa się jednakże bliżej sposobów ich stosowania ani nie podaje się kryteriów pozwalających na odróżnienie podstawowego i zaawansowanego ich wykorzystania. Zamierzeniem autora niniejszego tekstu jest wypełnienie tego niewątpliwego braku, a poniższe rozważania można zaliczyć do teoretyczno-metodologicznych analiz przetwarzania informacji. Są one często pomijane w literaturze przedmiotu, ponieważ

¹ Operatory te nazywa się operatorami Boole’a (boole’owskimi) lub operatorami logicznymi. Ze względów stylistycznych w niniejszym tekście używany będzie drugi termin, choć ściśle rzecz biorąc, są to operatory algebraiczne.

główny nacisk kładzie się na kwestie techniczne, aplikacyjne lub biznesowe [Unold, 2011, s. 393]. Stąd też zagadnienia, o których traktuje niniejszy tekst, mają charakter podstawowy i uznano je za wartość podjęcia. Co więcej, pomimo dynamicznego rozwoju narzędzi wyszukiwawczych, które mają być na tyle łatwe w użyciu, że nie wymagają stosowania operatorów logicznych, podkreśla się, że ich używanie wciąż jest i będzie wartościowym elementem strategii wyszukiwawczych. Pozwalają one na precyzyjne formułowanie zapytań wyszukiwawczych oraz umożliwiają użytkownikowi kontrolowanie ich formy i treści [Tortorici, 2015].

Podstawowe stosowanie operatorów logicznych

W wyszukiwarkach internetowych wykorzystuje się trzy operatory logiczne, które mają swoje odpowiedniki w algebrze zbiorów. Te operatory to: iloczyn (AND), suma (OR) i różnica (AND NOT). W wyniku wykorzystania operatora iloczynu do dwóch słów kluczowych otrzymujemy wyniki zawierające obydwa słowa kluczowe (jedno i drugie), następstwem zastosowania operatora sumy są wyniki zawierające dowolne (jedno lub drugie) słowo kluczowe, natomiast stosując operator różnicy, otrzymujemy wyniki zawierające pierwsze słowo kluczowe, ale nie zawierające drugiego z nich.

W wielu wyszukiwarkach, np. w Google i Bing, dla wykonywanych operacji można używać znaków symbolicznych: ‘|’ dla sumy oraz znaku ‘-’, umieszczonego bezpośrednio (bez spacji) przed wykluczonym słowem kluczowym, dla różnicy. W niniejszym tekście również będziemy symbolikę tę wykorzystywać. Ponadto, dla operatora iloczynu używać będziemy, wykorzystywanego w wyszukiwarce Bing, symbolu ‘&’. Ogólna charakterystyka operatorów jest taka, że służą one do *zawężania* wyników – iloczyn i różnica, lub do ich *rozszerzania* – suma.

Warto chwilę zatrzymać się nad charakterystyką powyższych operatorów. Po pierwsze, warto przypomnieć, że operator sumy nie ma charakteru rozłącznego, jak się niekiedy błędnie podaje². Jest to pomylenie operacji

² Można trafić na błędne informacje głoszące, że tak jest: „Znak potoku "|" (OR) – w konsekwencji zastosowania operatora uzyskamy wyniki uwzględniające zarówno jedno,

sumy (OR) z operacją różnicy symetrycznej (XOR) w algebrze zbiorów³. Po drugie, wbrew pozorom, operator różnicy nie jest operatorem jednoargumentowym. W algebrze zbiorów jest to pomylenie dwuargumentowej operacji różnicy (AND NOT) z jednoargumentową operacją dopełnienia zbioru (NOT), odpowiadającej negacji w logice. Takie pomylenie sugeruje sposób zapisu, wstawianie znaku '-' bez spacji przed niepożądanym słowem kluczowym. Gdyby był to operator jednoargumentowy, to możliwe byłoby wyszukiwanie czysto negatywne, wyszukiwanie tylko po wykluczonych słowach kluczowych, np. -„sport”. Jest to zadanie niewykonalne. Do operatorów logicznych zalicza się także cudzysłów oraz nawiasy [Lintz, Walters, 2015]. Choć, ściśle rzecz biorąc, nie są one operatorami logicznymi, to należy je traktować jako istotne elementy składni zapytań, analogicznie do znaków technicznych w logice. Stąd też w niniejszych rozważaniach również będziemy je wykorzystywać.

Weźmy teraz pod uwagę operator iloczynu. W większości wyszukiwarek, przede wszystkim w Google, jest on *operatorem domyślnym*. W tym sensie nawet początkujący użytkownicy wyszukiwarek internetowych stosują ten operator logiczny, choć być może nieświadomie [Hock, 2007, s. 70]. Jednak uznanie tego za stosowanie operatora logicznego trywializuje sprawę. Jak zatem rozumieć zamierzone użycie operatora iloczynu? Pewnym rozwiązaniem wydaje się przyjęcie, że polega ono na wstawieniu odpowiedniego symbolu. Byłoby to rozwiązanie dobre, gdyby nie fakt zmienności tej symboliki. Nie chodzi o to, że w różnych wyszukiwarkach stosuje się odmienne symbole, na przykład w Bing operator iloczynu jest domyślny, ale można także stosować symbole 'AND' lub '&'. W Google można było używać znaków 'AND' lub '+'. Symbole te były jednak stopniowo wycofywane, najpierw '+' (w roku 2011), a potem 'AND' (dwa lata później) [Schwarz, 2011; Shamaeva, 2013]. Przy domyślnym ustawieniu operatora iloczynu znak dla tej operacji jest redundantny. Pomimo to,

jak i drugie słowo, lecz nie występujące jednocześnie na tej samej stronie. Oznacza to, że wyszukiwarka nie będzie brała pod uwagę stron, na których obydwa słowa będą występować” [Kołacz, 2008].

³ Różnica symetryczna odpowiada alternatywnie rozłącznej w logice, suma – alternatywie nierozłącznej.

z pragmatycznego punktu widzenia, bywa on pomocny w precyzyjnym formułowaniu złożonych zapytań. Jego stosowanie byłoby wystarczającym (choć niekoniecznym) kryterium zamierzonego stosowania operatora iloczynu, jednak w Google już po prostu nie ma takiej możliwości. Z tego względu można mówić o pewnym spontanicznym lub nieświadomym stosowaniu tego operatora oraz stosowaniu zamierzonym. To pierwsze może być jednak błędne, a wtedy trudno w ogóle mówić o stosowaniu operatora logicznego. Jest to możliwe w sytuacji, gdy ktoś wpisuje dwa słowa kluczowe, w sposób nieświadomy realizuje domyślną operację iloczynu, gdy tymczasem z perspektywy potrzeb informacyjnych tej osoby należałoby użyć operatora sumy (np. „wakacje we Włoszech i Grecji”, zamiast „wakacje we Włoszech lub Grecji”) [Babik, 2010, s. 188]. Podanie pełnego i intersubiektywnego kryterium zamierzonego stosowania operatora iloczynu nie jest zatem sprawą łatwą. Z perspektywy problematyki niniejszego tekstu wystarczające będzie przyjęcie pewnego kryterium cząstkowego: ktoś w sposób zamierzony używa operatora iloczynu, jeśli potrafi go zastosować w kombinacji z innymi operatorami.

Stosowanie podstawowe proste operatorów logicznych, to pojedyncze zastosowanie operatora sumy lub operatora różnicy do dwóch słów kluczowych. To, czy stosowanie podstawowe obejmuje również spontaniczne użycie operatora iloczynu, gdy jest on operatorem domyślnym, pozostawiamy sprawą otwartą. Wydaje się jednak, że dobrym rozwiązaniem jest przyjęcie stosowania *operatorów niedomyślnych*.

Stosowanie podstawowe złożone to użycie więcej niż dwóch słów kluczowych i więcej niż jednego operatora. Obejmuje ono następujące sytuacje:

- iteracje operatora sumy;
- iteracje operatora różnicy;
- kombinacje iloczynu i różnicy.

Przykłady powyższych sytuacji:

- „komunikacja” | „łączność” | „transport”;
- „komunikacja” -, „łączność” -, „transport”;
- „komunikacja” & „łączność” -, „transport”.

Powtarzanie identycznych operatorów traktujemy jako stosowanie podstawowe. Jest to procedura łatwa w realizacji, a formułowane zapytania

mają jednoznaczny charakter. Nie ma tu zatem niebezpieczeństwa popełnienia błędu *amfibolii* – niejednoznaczności składniowej. W wyszukiwarkach, w których stosuje się nawiasy, wyszukiwania takie nie wymagają ich użycia. Pierwsze dwa przypadki są dość oczywiste. Jest to rozszerzanie wyników wyszukiwania poprzez powtarzanie operatora sumy i dodawanie słów kluczowych lub odpowiednie ich zawężanie poprzez eliminowanie kolejnych wyrażeń operatorem różnicy. Należy jednak zwrócić uwagę na kolejność wpisywanych słów kluczowych. Słowa te są rangowane z kolejnością ich wpisywania, od lewej do prawej, stąd też najlepiej wypisywać je w porządku ich relewancji. Dla jednowyrazowych słów kluczowych zamiana ich kolejności często nie ma znacznego wpływu na otrzymane wyniki, lecz w przypadku słów kluczowych wielowyrazowych ich kolejność bywa bardzo istotna, a wyniki wyszukiwań bardzo odmienne od siebie [Unold, 2011, s. 389–391]. Ważnym elementem procesu wyszukiwania jest także stosowanie cudzysłowu, także do jednowyrazowych słów kluczowych. Używanie cudzysłowu wraz z operatorami logicznymi znacznie zwiększa precyzję i relewancję wyników wyszukiwania [Unold, 2011, s. 398].

Sytuacja trzecia wymaga pewnego komentarza, ponieważ mamy tam kombinację dwóch operatorów – iloczynu i różnicy. Wyszukiwanie to ma jednak jednoznaczny charakter, chodzi o wyniki zawierające dwa słowa kluczowe, ale nie zawierające trzeciego. W wyszukiwarkach wykorzystujących nawiasy nie są one również potrzebne, ponieważ mogą być wtedy wstawione w dowolnym miejscu⁴. W Bing, gdzie nawiasy się stosuje, poniższe wyszukiwania⁵ dały identyczne wyniki (po 76 500):

- („komunikacja” & „łączość”) - „transport”;
- „komunikacja” & („łączość” - „transport”);
- „komunikacja” & „łączość” - „transport”.

Warto zaznaczyć jednak, że należy zachować pewien umiar w stosowaniu zbyt wielu operatorów jednocześnie, w szczególności nie należy

⁴ Koresponduje to z prawem algebry zbiorów: $A \cap (B - C) = (A \cap B) - C$. Nazywa się je prawem rozdzielnosci iloczynu względem różnicy [Kuratowski, Mostowski, 1966, s. 22].

⁵ Wszystkie kwerendy z dnia 17.04.2015 r.

nadmiernie multiplikować operatora sumy. Jeśli mamy długą listę słów kluczowych i jesteśmy zainteresowani wynikami, które zawierają niektóre z nich, wyszukiwanie za pomocą wielokrotnego zastosowania operatora sumy może nie przynieść oczekiwanych rezultatów [Shamaeva, 2014]. Lepszym rozwiązaniem może być wykorzystanie kilku wyszukiwań prostszych.

Podsumowując tę część rozważań, o stosowaniu podstawowym operatorów logicznych mówimy wtedy, gdy nie ma ryzyka popełnienia błędu amfibolii. Są to wyszukiwania czysto rozszerzające – stosowanie tylko operatora sumy, albo czysto zawężające – stosowanie tylko operatora różnicy lub kombinowanie operatora różnicy z iloczynem. Stosowanie podstawowe jest względnie niezależne od wykorzystanych wyszukiwarek, są to po prostu pewne elementarne strategie wyszukiwawcze wykorzystujące operatory logiczne. Stosowanie zaawansowane natomiast jest w silnym stopniu zależne od wykorzystanych serwisów wyszukiwawczych.

Zaawansowane stosowanie operatorów logicznych

Zaawansowane stosowanie operatorów logicznych obejmuje trzy sytuacje:

- kombinacje sumy i iloczynu;
- kombinacje sumy i różnicy;
- kombinacje wszystkich trzech operatorów.

Przy użyciu powyższych kombinacji zachodzi ryzyko popełnienia błędu amfibolii⁶. Są to także wyszukiwania, które nie mają charakteru czysto rozszerzającego albo czysto zawężającego, lecz charakter hybrydowy – *rozszerzająco-zawężający*. W dalszych partiach niniejszego tekstu zajmujemy się pierwszymi dwiema sytuacjami, ponieważ mają one charakter zasadniczy. Stosowanie kombinacji bardziej złożonych jest wynikiem ustalenia pewnych kombinacji bazowych. Zwraca się także uwagę na to, że

⁶ O błędzie amfibolii mówimy wtedy, gdy osoba formułująca niejednoznaczne składniowo zapytanie nie zdaje sobie z tej wieloznaczności sprawy [Ziemiński, 1994, s. 141].

należy zachować ostrożność przy wyszukiwaniu zawierającym wiele kombinacji nawiasowych (w wyszukiwarkach, w których nawiasy się stosuje) [Unold, 2011, s. 399–400]. Można przyjąć pragmatyczne zalecenie głoszące, że im więcej nawiasów, tym wyniki wyszukiwania mogą być mniej trafne. Z racji tego w dalszych partiach zajmujemy się kombinacjami jednonawiasowymi, które mają zasadniczy charakter. Dla większej przejrzystości ograniczymy się do wyszukiwania z trzema jednowyrazowymi słowami kluczowymi.

Zaawansowane zastosowania operatorów są mocno zależne od użytych serwisów wyszukiwawczych. Weźmy wyszukiwarki Google i Bing. Wyszukiwanie o treści „dom lub mieszkanie i garaż” jest dwuznaczne. Możemy mieć na myśli po prostu dom lub mieszkanie, ale z garażem, albo dom z garażem lub mieszkanie z garażem. Zapytanie to wpisane w Google zinterpretowane będzie na drugi sposób – „(dom lub mieszkanie) i garaż”, natomiast w Bing na sposób pierwszy – „dom lub (mieszkanie i garaż)” [Schamaeva, 2013a]. Jest to związane z odmienną kolejnością wykonywania operacji.

W Bing operatory mają różną moc – najpierw wykonywana jest operacja iloczynu, a następnie sumy. Google natomiast nie określa, czy w pierwszej kolejności powinien być wykonany operator sumy, czy iloczynu. W szczególności istotne to jest w przypadku operatora sumy. W Google działa on w taki sposób, że słowa kluczowe bezpośrednio przed i po nim występujące są automatycznie traktowane jako jego argumenty. W ten sposób zapytanie o formie:

A & B | C & D

jest identyczne z

A & (B | C) & D

[Hock, 2007, s. 71; Page, 2012].

Google nie rozpoznaje nawiasów, ich stosowanie nie ma wpływu na wyniki wyszukiwania. Wniosek z tego jest taki, że w Google argumentami operatora sumy mogą być tylko słowa kluczowe, lecz nie wyrażenia złożone, które zawierają inne operatory logiczne. W Bing stosuje się nawiasy,

więc aby sformułować zapytanie o formie $A \& (B | C) \& D$, należy nawisy wprowadzić, bez nich będzie to zapytanie o formie $(A \& B) | (C \& D)$. Pod względem stosowania operatorów logicznych Bing jest wyszukiwarką bardziej zaawansowaną. Ustala się tam kolejność wykonywanych operacji oraz stosuje się nawiasy. Daje to możliwość precyzyjnego formułowania zapytań wyszukiwawczych. Ogólna kolejność wykonywania operacji jest następująca: nawiasy, cudzysłów, różnica, iloczyn, suma⁷. Powyższe okoliczności mają decydujący wpływ na ustalenie zasadniczych kombinacji operatorów przydatnych w wyszukiwaniu zaawansowanym⁸.

Z logicznego punktu widzenia ważne są następujące kombinacje operatorów:

- sumy i iloczynu:
 - $A \& (B | C)$
 - $A | (B \& C)$
- sumy i różnicy:
 - $(A | B) - C$
 - $A | (B - C)$

Z powyższych dwóch par kombinacji szczególnie istotną funkcję pełni pierwsze z nich: $A \& (B | C)$ oraz $(A | B) - C$. W rozważaniach dotyczących stosowania operatorów logicznych są one wymieniane jako najważniejsze kombinacje operatorów w zapytaniach wyszukiwawczych i jako przykład dużej ich użyteczności [Bates, 2004; Lintz, Walters, 2015]. Są to te kombinacje, gdzie argumentami operatora sumy są słowa kluczowe, a nie wyrażenia złożone zawierające inny operator. Logicznie rzecz biorąc, są to kombinacje, w których suma nie jest *operatorem głównym*, lecz jest *operatorem zagnieżdżonym*. W pierwszej z wymienionych kombinacji głównym operatorem jest iloczyn, w drugiej – różnica. W Bing użycie tych kombinacji wymaga stosowania nawiasów, w pozostałych dwóch

⁷ Zasady działania operatorów logicznych są opisane w dziale pomocy dla wyszukiwarki Bing.

⁸ Konieczność stosowania nawiasów w zapytaniach wyszukiwawczych jest sprawą kontrowersyjną. Zwraca się uwagę na to, że ich stosowanie może wydawać się nieintuicyjne dla użytkowników zasobów informacyjnych, nawet dla tych, od których należałoby wymagać pewnych kompetencji w zakresie logiki. Dotyczy to np. prawników [Abramowicz, 2008, s. 226].

kombinacjach są one niekonieczne. Wymienione kombinacje są też możliwe do stosowania w Google. Pozostałych z nich w Google bezpośrednio wpisać nie można. To sprawia, że $A \& (B | C)$ oraz $(A | B) -C$ są kombinacjami wyróżnionymi. Są one jednak także wyróżnione z praktycznego punktu widzenia. Cechuje je pewna intuicyjna zrozumiałość i praktyczna przydatność, co wyjaśni się bliżej na przykładach. Pozostałe kombinacje, w których suma jest operatorem głównym, a jego argumentem wyrażenie zawierające inny operator, wydają się być mniej przydatne praktycznie. To, że suma jest operatorem głównym, sprawia też, że takie wyszukiwania zaawansowane łatwo mogą być zastąpione dwoma wyszukiwaniami prostymi:

- wyszukiwanie o formie $A | (B \& C)$ może być zastąpione osobnym wyszukiwaniem dla A i osobnym dla $B \& C$;
- wyszukiwanie o formie $A | (B -C)$ może być zastąpione osobnym wyszukiwaniem dla A i osobnym dla $B -C$.

Zapewne będzie tak, że w wynikach dwóch wyszukiwań prostych będą pojawiały się wyniki zdublowane. Jednak z formalnego punktu widzenia są to dwa osobne wyszukiwania, ponieważ słowa kluczowe użyte w obu z nich się nie powtarzają. Do wątku tego powrócimy w końcowej części niniejszego artykułu.

Jak stwierdzono powyżej, kombinacje $A \& (B | C)$ oraz $(A | B) -C$, oczywiście bez nawiasów i znaku iloczynu, mogą być stosowane w Google. O tym, że one „działają”, świadczy fakt, że cechuje je pewna stabilność, którą formalnie można wyrazić jako:

- $[A \& (B | C)] \cong [(B | C) \& A]$
- $[(A | B) -C] \cong [-C (A | B)]$

Sprowadza się to do tego, że wyszukiwania po zamianie miejscami argumentów operatora głównego, przynajmniej dla jednowyrazowych słów kluczowych ujętych w cudzysłów, powinny dawać identyczne lub zbliżone wyniki. Poniżej przedstawione przykłady wyszukiwań pokazują, że w Google tak rozumiana stabilność zachodzi. Nie dotyczy to jednak pozostałych dwóch kombinacji, co dodatkowo wyróżnia wybrane kombinacje operatorów. W Bing stabilność ta jest zapewniona stosowaniem nawiasów.

Warto w tym miejscu zwrócić uwagę na ciekawy fakt, który dotyczy operatora różnicy. Otóż, zgodnie z drugim z powyższych warunków stabilności kombinacji operatorów, słowo kluczowe poprzedzone operatorem różnicy może stać na początku zapytania wyszukiwawczego. Z praktycznego punktu widzenia jest to dość nieintuicyjne, a z logicznego punktu widzenia – składniowo niepoprawne⁹. Niemniej jednak, jak pokazują poniższe przykłady kwerend, dla jednowyrazowych słów kluczowych ten warunek stabilności zachodzi. Pokazuje to specyfikę operatora różnicy – choć jest to operator dwuargumentowy, to w pewien sposób „zachowuje się” jak jednoargumentowy. Natomiast pierwszy z warunków stabilności dla jednowyrazowych słów kluczowych ujętych w cudzysłów raczej nie wzbudza zastrzeżeń.

Weźmy kombinację **A & (B | C)** i założmy, że ktoś w Google poszukuje informacji na temat logiki, ale nie tylko samej logiki, lecz wraz z informacjami na temat semiotyki lub retoryki. Otrzymujemy następujące wyniki:

- „logika” „semiotyka” | „retoryka” – 121 000;
- „semiotyka” | „retoryka” „logika” – 121 000.

Dla porównania:

- „logika” „semiotyka” – 40 100;
- „logika” „retoryka” – 109 000;
- „semiotyka” | „retoryka” – 403 000.

Inne wyszukiwanie – w Google poszukujemy informacji na temat filozofii, ale wraz z informacjami na temat logiki lub metodologii:

- „filozofia” „logika” | „metodologia” – 474 000;
- „logika” | „metodologia” „filozofia” – 474 000;
- „filozofia” „logika” – 425 000;
- „filozofia” „metodologia” – 206 000;
- „logika” | „metodologia” 44 700 000.

Powyższe wyniki ilustrują rozszerzająco-zawężający charakter wyszukiwań opartych na kombinacji **A & (B | C)**. Rozszerza ona wyniki w sto-

⁹ W algebrze zbiorów operator różnicy jest nieprzemiennej: $(A - B) \neq (B - A)$. Natomiast w kombinacji $(A \cup B) - C$, nie można zamienić argumentów miejscami. Wyrażenie o formie: $- C (A \cup B)$ jest syntaktycznie nieprawne.

sunku do wyszukiwań $A \& B$ oraz $A \& C$ branych z osobna, natomiast zawęża je w stosunku do wyszukiwania $B | C$. Zachodzi zatem zakresowa zgodność oraz wspomniana powyżej stabilność. Rozszerzająco-zawężający charakter tych wyszukiwań sprawia, że kombinacja $A \& (B | C)$ *optymalizuje* procesy wyszukiwania informacji. Nie jest to tylko rozszerzanie albo tylko zawężanie wyników. To samo dotyczy drugiej z wyróżnionych kombinacji.

Weźmy teraz kombinację $(A | B) - C$ i założmy, że ktoś w Google poszukuje informacji na temat komunikacji lub łączności, ale nie na temat transportu:

- „komunikacja” | „łączność” - „transport” – 22 400 000;
- - „transport” „komunikacja” | „łączność” – 22 400 000.

Dla porównania:

- „komunikacja” - „transport” – 17 900 000;
- „łączność” - „transport” – 468 000;
- „komunikacja” | „łączność” – 27 900 000.

Inne wyszukiwanie – w Google poszukujemy informacji o retoryce lub semiotyce, ale tym razem bez informacji o logice:

- „retoryka” | „semiotyka” - „logika” – 383 000;
- - „logika” „retoryka” | „semiotyka” – 383 000;
- „retoryka” - „logika” – 351 000;
- „semiotyka” - „logika” – 45 000;
- „retoryka” | „semiotyka” – 403 000.

Również mamy tu przypadek optymalizacji procesu wyszukiwania informacji. Kombinacja $(A | B) - C$ rozszerza wyniki w stosunku do wyszukiwań $A - C$ oraz $B - C$ branych z osobna, natomiast zawęża je w stosunku do wyszukiwania $A | B$. Mamy tu także ilustrację stabilności tej kombinacji i specyficznej *quasi-przemienności* operatora różnicy.

Oprócz tego, że powyższe kombinacje optymalizują procesy wyszukiwania informacji, to ponadto są one względem siebie *komplementarne*. Celem zastosowania pierwszej z nich jest uzyskanie wyników z takiego obszaru, który przez zastosowanie drugiej z nich ma być wykluczony i odwrotnie. Weźmy następujące słowa kluczowe: A – „semiotyka”, B – „retoryka” i C – „logika”, wariant kombinacji pierwszej $(A | B) \& C$ oraz

kombinację drugą ($A \mid B$)-C. W wyniku użycia kombinacji pierwszej otrzymujemy informacje o logice wraz z informacjami o semiotyce lub retoryce. Informacje te są natomiast wykluczone przez użycie kombinacji drugiej. Celem jej zastosowania jest bowiem uzyskanie informacji o semiotyce lub retoryce, ale bez informacji o logice. Komplementarność ta dodatkowo potwierdza to, że w zapytaniach wyszukiwawczych kombinacje te pełnią kluczową rolę.

Dla pełności rozważań weźmy pod uwagę pozostałe kombinacje: $A \mid (B - C)$ oraz $A \mid (B \& C)$. W Bing są to kombinacje stabilne w powyżej przyjętym sensie ze względu na to, że suma jest operatorem o najniższym priorytecie. Nie jest nawet wówczas konieczne stosowanie nawiasów. W Google kombinacje te nie są stabilne, zamiana miejscami argumentów operatora sumy, gdy jest on operatorem głównym, daje bardzo rozbieżne wyniki.

Dla kombinacji $A \mid (B - C)$:

- „komunikacja” \mid „łączność” - „transport” – 22 500 000;
- „łączność” - „transport” \mid „komunikacja” – 458 000.

Dla kombinacji $A \mid (B \& C)$:

- „retoryka” \mid „semiotyka” „logika” – 121 000;
- „semiotyka” „logika” \mid „retoryka” – 42 500.

Rekapitulując, możemy wskazać dwie bardzo przydatne jednonawiasowe kombinacje operatorów. Charakterystyczne dla nich jest to, że operator sumy jest zagnieżdżony w nawiasie. Pełnią one rolę rozszerzająco-zawężającą w odniesieniu do wyników, a także skracającą w odniesieniu do zapisu. Jeśli bowiem ktoś poszukuje informacji o logice, ale wraz z informacjami o semiotyce lub retoryce, to znaczy, że poszukuje informacji o logice i semiotyce lub o logice i retoryce. Zaawansowane wyszukiwanie o formie $A \& (B \mid C)$, może być zatem zastąpione dwoma wyszukiwaniami prostymi: $A \& B$ oraz $A \& C$ ¹⁰. Natomiast jeśli ktoś poszukuje informacji o komunikacji lub łączności, ale nie o transporcie, to znaczy, że poszukuje informacji o komunikacji, ale nie o transporcie lub o łączności, ale nie

¹⁰ Koresponduje to z prawem rozdzielności iloczynu względem sumy: $(A \cap B) \cup (A \cap C) = A \cap (B \cup C)$, które w zastosowaniach praktycznych pozwala na eliminację elementów redundantnych oraz pełni funkcję skracającą [Siegfried, 1977, s. 33].

o transporcie. Wyszukiwanie o formie $(A | B) - C$, może być więc rozłożone na dwa wyszukiwania proste: $A - C$ oraz $B - C$ ¹¹. Kombinacje te użyte w jednym wyszukiwaniu są logicznie zastępowalne dwoma osobnymi wyszukiwaniami: o logice i semiotyce oraz o logice i retoryce z osobna; o komunikacji, ale nie transporcie i o łączności, ale nie transporcie z osobna. Użycie tych kombinacji pozwala jednak w jednym wyszukiwaniu wyeliminować sporą liczbę wyników redundantnych. Biorąc się one stąd, że w obydwu wyszukiwaniach prostych powtarzają się słowa kluczowe. W szczególności dotyczy to pierwszej sytuacji, gdzie w dwóch wyszukiwaniach prostych powtarza się słowo kluczowe A . Odwołując się do przywołanych powyżej przykładów:

- „logika” „semiotyka” | „retoryka” – 121 000;
- „logika” „semiotyka” + „logika” „retoryka” – 149 100;
- „filozofia” „logika” | „metodologia” – 474 000;
- „filozofia” „logika” + „filozofia” „metodologia” – 631 000.

Z dużym prawdopodobieństwem można wnosić, że różnica pomiędzy sumą wyników dwóch wyszukiwań prostych a ilością wyników jednego wyszukiwania zaawansowanego jest ilością wyników redundantnych (zdubowanych). W przypadku pierwszej pary mamy redukcję ilości wyników o około 19%, w przypadku drugiej – o około 25%.

W ogólności użyteczność wyszukiwawczą kombinacji $A \& (B | C)$ można wyrazić następująco:

$$i [A \& (B | C)] < [i (A \& B) + i (A \& C)],$$

gdzie przez i rozumiemy ilość wyników wyszukiwania. Stwierdza się tutaj, że ilość wyników wyszukiwania zaawansowanego powinna być mniejsza od sumy ilości wyników dwóch wyszukiwań prostych. Oprócz tego użyteczność wyszukiwawcza tej kombinacji przejawia się także w tym, że ilość wyników dzięki niej uzyskana powinna być większa od ilości wyników zwracanych przez odpowiadające jej wyszukiwania proste branych z osobna:

$$i [A \& (B | C)] > i (A \& B);$$

$$i [A \& (B | C)] > i (A \& C).$$

¹¹ Koresponduje to z prawem $(A - C) \cup (B - C) = (A \cup B) - C$.

Ustalenie analogicznej zależności pierwszego rodzaju dla kombinacji $(A | B) - C$ jest kwestią raczej wątpliwą, ponieważ powtarzające się słowo kluczowe w wyszukiwaniach prostych jest poprzedzone operatorem różnicy, czyli jest to słowo kluczowe negatywne. W poszczególnych przypadkach zależności pomiędzy wynikami wyszukiwania będą odmiennie się kształtować. Niemniej jednak powinny być spełnione zależności rodzaju drugiego:

$$i [(A | B) - C] > i (A - C);$$

$$i [(A | B) - C] > i (B - C).$$

Zakończenie

W artykule zaproponowano odróżnienie podstawowego i zaawansowanego stosowania operatorów logicznych w zapytaniach wyszukiwawczych. Można wskazać trzy kryteria takiego odróżnienia. Pierwsze z nich ma charakter syntaktyczny – w przypadku użycia podstawowego nie ma ryzyka popełnienia błędu amfibolii, które zachodzi w przypadku użycia zaawansowanego. Drugie kryterium ma charakter w znacznej mierze semantyczny, dotyczy ono stosunków zakresowych zachodzących pomiędzy zbiorami wyników wyszukiwań. Stosowanie podstawowe to wyszukiwania tylko rozszerzające albo tylko zawężające. Stosowanie zaawansowane to wyszukiwania rozszerzająco-zawężające. Kryterium trzecie ma charakter pragmatyczny i uwzględnia aspekt technologiczny. Użycie proste jest względnie niezależne od użytych serwisów wyszukiwawczych, użycie zaawansowane – jest wobec nich zależne w wysokim stopniu.

Wyróżniono także kluczowe kombinacje charakterystyczne dla wyszukiwania zaawansowanego. Kombinacje te są komplementarne w tym sensie, że przy ustalonych wspólnych słowach kluczowych pierwsza kombinacja daje takie wyniki, które druga wyklucza, i odwrotnie. Możemy także powiązać z nimi charakterystyczne postawy wyrażające preferencje informacyjne. Dla $A \& (B | C)$ będzie to: „Poszukuję informacji na temat A, ale

wraz z informacjami na tematy B lub C” oraz „Poszukuję informacji na tematy A lub B, ale nie interesują mnie informacje na temat C”, dla $(A \mid B) - C$.

Analizy zawarte w niniejszym artykule oraz uzyskane wyniki opierają się na przyjęciu trzech upraszczających założeń:

- uwzględnienie tylko operatorów *stricte* logicznych;
- ograniczenie się do jednowyrazowych słów kluczowych ujętych w cudzysłów;
- wyróżnienie tylko kombinacji jednonawiasowych.

Założenia te nie mają jednak charakteru arbitralnego. Po pierwsze, ich przyjęcie jest motywowane potrzebą rozpatrzenia pewnych sytuacji elementarnych, co pozwala na ustalenie zależności mających charakter zasadniczy. Po drugie, wyniki wyszukiwań oparte na ustalonych zależnościach cechuje wysoka relewancja. Rozpatrzenie sytuacji i zależności, dla których powyższe założenia są uchylone, jest sprawą osobnych studiów.

Bibliografia

- Abramowicz W., (2008), *Filtrowanie informacji*, Poznań, Wydawnictwo Akademii Ekonomicznej w Poznaniu.
- Babik W., (2010), *Słowa kluczowe*, Kraków, Wydawnictwo Uniwersytetu Jagiellońskiego.
- Bates M.E., (2004), “Is Boolean dead?”, *EContent*, nr 4/04, www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=6268&ContextSubtypeID=13 [dostęp: 02.05.2015].
- BING pomoc, *Zawansowane opcje wyszukiwania*, onlinehelp.microsoft.com/pl-pl/bing/ff808438.aspx [dostęp: 02.05.2015].
- Hock R., (2007), *The Extreme Searcher's Internet Handbook: a Guide for the Serious Searcher*, New Jersey, CyberAge Books.
- Katalog kompetencji medialnych, informacyjnych i cyfrowych*, (2014), edukacjamedialna.edu.pl/media/chunks/attachment/Katalog_kompetencji_medialnych_2014_online.pdf [dostęp: 02.05.2015].
- Kończak M., (2008), *Google – zaawansowane wyszukiwanie*, www.hotfix.pl/google-zaawansowane-wyszukiwanie--a60.htm [dostęp: 02.05.2015].
- Kuratowski K., Mostowski A., (1966), *Teoria mnogości*, Warszawa, PWN.
- Lintz A., Walters K., (2015), *Boolean is not an Alien Language*, vimeo.com/125258591 [dostęp: 02.05.2015].

- Page E., (2012), *Order of Operations on Google*, erinpagemd.wordpress.com/2012/11/28/order-of-operations-on-google [dostęp: 02.05.2015].
- Siegfried H.J., (1977), *Od teorii mnogości do algebry logiki. Zastosowanie algebry logiki w technice cyfrowej*, Warszawa, Wydawnictwa Komunikacji i Łączności.
- Schamaeva I., (2013a), *Live AND Learn (Important Clarifications on Search Syntax)*, booleanstrings.com/2013/01/03/live-and-learn-important-clarifications-on-search-syntax, [dostęp: 02.05.2015].
- Schamaeva I., (2013b), *There's No AND*, booleanstrings.com/2013/06/26/theres-no-and, [dostęp: 02.05.2015].
- Schamaeva I., (2014), *Stop Using Boolean OR on Google*, booleanstrings.com/2014/12/19/stop-using-boolean-or-on-google [dostęp: 02.05.2015].
- Schwarz B., (2011), *Google Removes The + Search Command*, searchengine-land.com/google-sunsets-search-operator-98189 [dostęp: 02.05.2015].
- Torticci M., (2015), *Boolean is Not Dead*, <http://www.sourcecon.com/news/2015/02/10/boolean-is-not-dead-by-marknexus> [dostęp: 02.05.2015].
- Unold J., (2011), *Teoretyczno-metodologiczne podstawy przetwarzania informacji w cyberprzestrzeni*, Wrocław, Wydawnictwo Uniwersytetu Ekonomicznego.
- Ziemiński Z., (1994), *Logika praktyczna*, Warszawa, PWN.

Boolean operators in search queries

ABSTRACT. This article deals with using Boolean operators in search engine queries. The author distinguishes between the basic and the advanced use of these operators based on three criteria. The first is a syntactic one. As for basic use cases, there is no risk of an amphiboly error, which occurs with advanced use. The second criterion is a semantic one. It concerns the relationships between sets of search results. The basic use of operators either only expands these results or only restricts them. On the other hand, advanced usage both expands and restricts search results at the same time. The last criterion is a pragmatic one and it takes into account technological aspects. Basic use is relatively independent of the search engines that are being utilized, while advanced use is highly dependent on the search engines in use. The author also draws attention to two combinations of operators that are used in advanced search. They are very useful for formulating complex and non-trivial queries because they allow one to significantly reduce the number of redundant search results.

KEY WORDS: search engines, search queries, information literacy, Boolean operators, amphiboly

Jarosław Boruszewski, Zakład Teorii i Filozofii Komunikacji, Instytut Filozofii, Uniwersytet im. Adama Mickiewicza, ul. Szamarzewskiego 89C, 60-569 Poznań, borjar@amu.edu.pl