



Opracowanie przedmiotowe dokumentów
z zakresu nauk ścisłych: matematyczno-przyrodniczych i technicznych.
Język haseł przedmiotowych KABA:
teoria, praktyka, przyszłość

Kazimierz Dolny, 20-22 września 2006

[Poprzedni](#) - [Spis treści](#) - [Następny](#)

Andrzej Krajka
Artur Ściborek

Instytut Bibliotekoznawstwa i Informacji Naukowej. Uniwersytet im. Marii Curie-Skłodowskiej,
Lublin
akrajka@klio.umcs.lublin.pl

Ośrodek Informatyki Akademii Rolniczej w Lublinie
artur@ar.lublin.pl

Schematy zachowań użytkowników języka haseł przedmiotowych KABA

Patterns of behaviour of the SIS KABA language users

Słowa kluczowe

Jhp KABA, wyszukiwanie informacji, logi transakcyjne, katalogi OPAC, przeformułowanie zapytań, sesja użytkownika, wątki wyszukiwawcze

Keywords

Subject Indexing System KABA, information retrieval, transaction log analysis, OPAC Catalogues, reformulation of query, session of user, thread of retrieval,

Abstrakt

Na podstawie bogatego materiału badawczego logów transakcyjnych Biblioteki Głównej AR w Lublinie przeanalizowaliśmy różne sposoby wyszukiwania dokumentów ze szczególnym uwzględnieniem jhp KABA. Śledząc terminy, które użytkownik po kolei wprowadzał do katalogu OPAC, staraliśmy się przeanalizować jego tok rozumowania i wnikać możliwie głęboko w motywy podjętej strategii wyszukiwania informacji. Było to szalenie trudne, jak bowiem z kolejnych pytań zadawanych serwerowi można odtworzyć tok myślenia użytkownika? Często zadawaliśmy identyczne pytania, jakich używał użytkownik, aby zobaczyć, jakie były efekty takiego wyszukiwania. Analiza taka również wymagała szerokiej wiedzy dotyczącej zagadnień wyszukiwanych przez użytkowników, który mógł wyszukiwać praktycznie wszystko. Ponadto, zastanawiając się nad funkcjonowaniem całego interfejsu www katalogu komputerowego, opisaliśmy najczęstsze błędy popełniane przez użytkowników w trakcie wyszukiwania. Przebadane błędy pozwalają ulepszyć metody pracy katalogu komputerowego według oczekiwań użytkowników.

Abstract

On the basis of the rich research material i.e the analysis of the transaction logs of the Main Library of the Agricultural University in Lublin, the method of searching the KABA language has been analyzed and the searching have been compared with the other searching opportunities of the OPAC catalogue. The analyzing the retrieval terms, sequentially written by users, lead us on the way of reasoning of users (under the retrieval session) and penetrate the psychologically aspects of there retrieval strategy. It was very difficult. We often have searched identically as users in order to see these retrieval results. Furthermore, we must have had the wide knowledge which deals with these retrieval problems. Additionally, in order to investigate the users interface we will describe the errors made by the users. These errors allows us to improve user's interface.

1. MATERIAŁ ŹRÓDŁOWY

1.1. Logi transakcyjne katalogu bibliotecznego OPAC

Analiza rejestrów katalogu bibliotecznego OPAC jest to jedna z najważniejszych metod używanych przez badaczy-praktyków wyszukiwania informacji, która umożliwia analizę komunikacji użytkownika z systemem wyszukiującym (zob. [1], [6]). Choć metoda ta używana była w systemach bibliecznych i systemach wyszukiwania informacji (zob. [3],[4],[5]), stała się jednym z najbardziej odpowiednich sposobów zbierania danych od wielu końcowych użytkowników. Rejestry z serwera często zawierają adresy IP (lub nazwy domen) wyszukiwujących, datę i godzinę wyszukiwania, komendy wyszukiwania oraz wyniki wyszukiwania. Takie rejestry są użyteczne przy tworzeniu ogólnych statystyk na tematy poszukiwane, wyciągania wniosków dotyczących podjętej strategii wyszukiwawczej, a także testowania hipotez na temat efektów różnych funkcji systemu w oparciu o dane dotyczące zachowań[2]. W naszym badaniu analiza rejestru była potrzebna, gdyż zawiera on szczegółowe dane na temat modyfikacji zapytań. Materiał do badań uzyskaliśmy dzięki uprzejmości wicedyrektora Biblioteki Głównej Akademii Rolniczej w Lublinie. W dostarczonych nam logach transakcyjnych każdy rekord opisywał jedną akcję użytkownika. W tym rekordzie znajdowały się następujące informacje (zob. Rys. 1):

Adres IP użytkownika

Identyfikator sesji

Czas, w którym użytkownik dokonał transakcji (czas wciśnięcia klawisza ENTER)

Identyfikator ekranu

Terminy wyszukiwawcze lub numer wyszukiwanego rekordu bibliograficznego

Odpowiedź systemu (ile pozycji zostało wyświetlonych lub przesłanych)

Zarejestrowane wyszukiwania dotyczyły wielu tematów w różnych zapytaniach podczas jednej sesji z komputerowym katalogiem bibliotecznym. Ponieważ w dalszej analizie i na wykresach często dla ułatwienia stosować będziemy kody, dlatego pokrótce wyjaśniamy tutaj ich znaczenie. Wyjaśnijmy kody typu wyszukiwania z użyciem indeksów wyszukiwawczych:

1003 – wyszukiwanie w indeksie autorskim

21 - wyszukiwanie w indeksie haseł przedmiotowych – KABA

4 - wyszukiwanie w indeksie tytułowym

2019 - wyszukiwanie w indeksie tytułów czasopism

2009 - wyszukiwanie w indeksie haseł przedmiotowych w języku angielskim/francuskim –

KABA

20 - wyszukiwanie w indeksie sygnatur.

W przypadku dowolnych słów kluczowych można było używać kodów szukaj wszędzie lub wyszukiwanie w szczególnym typie publikacji, jak wydawnictwo zwarte, ciągłe, dysertacja/praca doktorska. Natomiast kody typów transakcji to:

1. SCAN (w skrócie S) - przeglądanie/wyszukiwanie haseł z dostępnych indeksów:
 - autorów - zawiera nazwy osób, instytucji lub organizacji
 - tytułów - zawiera tytuły książek, czasopism, norm, serii wydawniczych itp.
 - tytułów czasopism
 - haseł przedmiotowych - zawiera hasła przedmiotowe w języku haseł przedmiotowych KABA charakteryzujące treść dostępnych dzieł
 - haseł przedmiotowych w języku angielskim i francuskim
 - sygnatur - zawiera sygnatury książek i czasopism.
2. BIBREQ (w skrócie B) - wyświetlenie rekordu bibliograficznego
3. KW (w skrócie K) - wyszukiwanie przez słowa kluczowe - słowa kluczowe są to słowa zaczerpnięte z indeksowanych pól opisu bibliograficznego oraz z towarzyszących opisowi haseł.
4. KIH (w skrócie H) - wyszukiwanie przez słowa z haseł - jest to wyszukiwanie słów w hasłach występujących w indeksach: autorów, tytułów i haseł przedmiotowych.

Na przykład:

S21 – oznacza przeszukiwanie indeksu haseł przedmiotowych jhp KABA

S1003 – oznacza przeszukiwanie indeksu autorskiego

S4 – to przeszukiwanie indeksu tytułowego

K21 – oznacza przeszukiwanie przez słowa kluczowe z haseł jhp KABA

K4 – oznacza przeszukiwanie przez słowa kluczowe z haseł indeksu tytułowego

Dane pobrane zostały z rejestru zapytań za okres czternastu dni, tj. od 25 stycznia 2006 r. do 7 lutego 2006 r. Uzyskaliśmy w ten sposób 11 692 rekordów od użytkowników korzystających z terminali komputerowych zlokalizowanych w budynku Biblioteki Głównej, jak również z dowolnych komputerów podłączonych do Internetu. Dzięki temu, że w rejestrze zapisany był adres IP komputera, data, godzina, minuta, sekunda oraz ID użytkownika, czyli jednoznaczny identyfikator numeru sesji (nr sesji rozumiany jako otwarcie okna przeglądarki WWW z adresem katalogu OPAC), mogliśmy posegregować dane tak, by analizować zachowanie wyszukiwawcze i przeformułowania wielokrotne zapytań jednego użytkownika. Do tej analizy wybraliśmy tylko te sesje, w których użytkownicy formułowali więcej niż dwa zapytania. Analiza przeformułowań prowadzona była zarówno poprzez badanie dwóch kolejnych operacji wyszukiwania (tylko wyszukiwania!) użytkownika, jak i poprzez analizę całej sesji użytkownika. Do badania schematów zachowań wielokrotnych przeformułowań wykorzystano własne programy komputerowe napisane w języku Visual Basic for MS Access i MS Excel, dzięki którym strategia wyszukiwawcza i zachowania użytkownika stały się łatwiejsze do analizy. Poniższy przykład pokazuje użytkownika, który przeprowadzał wyszukiwanie przez 31 min. (14.52-15.23).

Adres IP komputera	Numer sesji	Data, czas	Typ transakcji	Zapytanie
83.7.14.61	2006012514512826587	01:25:2006:14:52:08	[SCAN]	1003\Fitosocjologia stosowana
83.7.14.61	2006012514512826587	01:25:2006:14:52:56	[SCAN]	21\Fitosocjologia
83.7.14.61	2006012514512826587	01:25:2006:14:53:31	[SCAN]	21\Fitosocjologia stosowana
83.7.14.61	2006012514512826587	01:25:2006:14:53:56	[SCAN]	4\Fitosocjologia stosowana
83.7.14.61	2006012514512826587	01:25:2006:14:54:04	[BIBREQ]	VtIs000040748
83.7.14.61	2006012514512826587	01:25:2006:14:54:43	[SCAN]	4\zdjęcia fitosocjologiczne
83.7.14.61	2006012514512826587	01:25:2006:14:55:27	[SCAN]	21\Fitosociologia
83.7.14.61	2006012514512826587	01:25:2006:14:57:43	[SCAN]	21\geobotanika
83.7.14.61	2006012514512826587	01:25:2006:14:58:02	[SCAN]	4\Geobotanika
83.7.14.61	2006012514512826587	01:25:2006:14:58:39	[SCAN]	21\zespolo roślinne
83.7.14.61	2006012514512826587	01:25:2006:14:59:03	[SCAN]	4\Zespoły roślinne
83.7.14.61	2006012514512826587	01:25:2006:14:59:11	[BIBREQ]	vtIs000004090
83.7.14.61	2006012514512826587	01:25:2006:14:59:41	[SCAN]	4\fitocenologia
83.7.14.61	2006012514512826587	01:25:2006:14:59:48	[BIBREQ]	vtIs000007118
83.7.14.61	2006012514512826587	01:25:2006:15:00:20	[SCAN]	21\fitocenologia
83.7.14.61	2006012514512826587	01:25:2006:15:03:11	[SCAN]	21\fitocenologia
83.7.14.61	2006012514512826587	01:25:2006:15:03:33	[SCAN]	4\Fitosociologia
83.7.14.61	2006012514512826587	01:25:2006:15:03:44	[BIBREQ]	vtIs000023118
83.7.14.61	2006012514512826587	01:25:2006:15:03:54	[BIBREQ]	vtIs000039337
83.7.14.61	2006012514512826587	01:25:2006:15:04:03	[BIBREQ]	vtIs000040748
83.7.14.61	2006012514512826587	01:25:2006:15:05:26	[SCAN]	21\Fitosocjologia
83.7.14.61	2006012514512826587	01:25:2006:15:05:42	[SCAN]	20\Fitosocjologia
83.7.14.61	2006012514512826587	01:25:2006:15:05:57	[SCAN]	2019\Fitosocjologia
83.7.14.61	2006012514512826587	01:25:2006:15:06:08	[SCAN]	4\Fitosocjologia
83.7.14.61	2006012514512826587	01:25:2006:15:06:53	[SCAN]	21\zestawy roślinności
83.7.14.61	2006012514512826587	01:25:2006:15:07:53	[SCAN]	21\rowy do nawodnień
83.7.14.61	2006012514512826587	01:25:2006:15:08:56	[SCAN]	21\rowy melioracyjne
83.7.14.61	2006012514512826587	01:25:2006:15:09:28	[SCAN]	21\spis florystyczny
83.7.14.61	2006012514512826587	01:25:2006:15:09:59	[SCAN]	1003\zestawy roślinności
83.7.14.61	2006012514512826587	01:25:2006:15:10:32	[SCAN]	21\jezioro bikcze
83.7.14.61	2006012514512826587	01:25:2006:15:12:39	[SCAN]	21\zbiorniki retencyjne
83.7.14.61	2006012514512826587	01:25:2006:15:13:34	[SCAN]	21\zbiorniki
83.7.14.61	2006012514512826587	01:25:2006:15:17:41	[SCAN]	21\zestawy roślin
83.7.14.61	2006012514512826587	01:25:2006:15:18:12	[SCAN]	21\zbiorniki roślin
83.7.14.61	2006012514512826587	01:25:2006:15:19:11	[SCAN]	1003\zbiorniki odwadniające
83.7.14.61	2006012514512826587	01:25:2006:15:19:30	[SCAN]	21\zbiorniki odwadniające
83.7.14.61	2006012514512826587	01:25:2006:15:19:58	[SCAN]	21\nawadnianie

83.7.14.61	2006012514512826587	01:25:2006:15:21:46	[SCAN]	21\;rowy nawadniająca
83.7.14.61	2006012514512826587	01:25:2006:15:22:07	[SCAN]	21\;nawadnianie
83.7.14.61	2006012514512826587	01:25:2006:15:22:16	[BIBREQ]	vtls000037955
83.7.14.61	2006012514512826587	01:25:2006:15:23:18	[SCAN]	1003\;odwadnianie

Rys. 1. Przykład jednoznacznie zidentyfikowanej sesji użytkownika

Widać na początku sesji pomyłkę - temat dotyczący *fitosocjologii stosowanej* (dyscyplina botaniczna, mająca za przedmiot badanie występujących w naturze zbiorowisk roślinnych) wyszukiwał użytkownik według indeksu autorskiego, potem poprawił się i zaczął wyszukiwać to słowo w indeksie haseł przedmiotowych, ale ogólniejsze hasło *fitosocjologia*. Po otrzymaniu przypuszczalnie zbyt dużej liczby luźno związanych odpowiedzi użytkownik uszczegółowił swoje pytanie, wyszukując w indeksie haseł przedmiotowych hasła *fitosocjologia stosowana*, następnie zastosował kolejno terminy uszczegóławiające, skojarzone, uogólnione, synonimy. Kolejnym tematem wyszukiwawczym stały się terminy związane z *jeziorami*. Również w tym przypadku użytkownik zastosował strategię wyszukiwawczą z użyciem uogólnień, skojarzeń, uszczegółowień.

1.2. Opis przeformułowań i wątków wyszukiwawczych

Podstawą do dalszej analizy była modyfikacja zapytań przez użytkownika. Ponieważ skupiliśmy się tu na przeformulowywaniu zapytań, nie analizowaliśmy wyświetleń ani jakości znalezionych rekordów bibliograficznych. Dane, które posiadaliśmy, pokazywały tylko strategię wyszukiwawczą oraz liczbę rekordów bibliograficznych, które użytkownik dzięki zastosowanej strategii odnalazł.

W pierwszym etapie obróbki otrzymanego materiału połączyliśmy kolejne operacje wyszukiwania użytkowników w ramach jednej sesji (w tabeli LONGLOG). Jednakże w obrębie sesji użytkownik mógł wyszukiwać jeden lub grupę związanych tematycznie dokumentów albo mógł realizować kilka czasem zupełnie różnych poszukiwań. Dlatego ważnym problemem stało się wyodrębnienie wątków wyszukiwawczych, czyli powiązanych tematycznie zagadnień, które wyszukiwał użytkownik.

Ogólne podejście do sesji w tabeli LONGLOG musieliśmy uzupełnić szczegółową analizą kolejnych zapytań użytkownika. Każde dwie kolejne operacje wyszukiwania w ramach tej samej sesji były analizowane z punktu widzenia kroków, jakie wykonał użytkownik. Przykładowy fragment otrzymanej tabeli (o nazwie TWOLOG) wraz z wyszczególnionymi kilkunastoma przykładami podstawowych typów przeformułowań przedstawiamy na Rysunku 2. Oprócz wymienionych podstawowych typów niektóre przeformułowania opisywano dowolną ilością kombinacji połączeń tych typów w danej sesji lub wątku.

Kwerenda nr 1	Kwerenda nr 2	Typ przeformułowania
1003\;Acta microbiologica hungarica	2019\;Acta microbiologica hungarica	a
1003\;Żukowska-Biemans	1003\;Żukowska-Biemans	b
2019\;asae	2019\;transitions asae	u
21\;technologie mięsa	21\;mięsa	o
21\;tlen aktywne formy	4\;aktywne formy tlenu	d
4\;zadrzewienia	4\;zadrzewianie	p
1003\;gwroński	1003\;gawroński	1
1035\;chodowla kur	1035\;chodowla kór	2
1003\;blasinski	1003\;pyc	e
4\;spożywcze	4\;spożywczym	h
1003\;Krzysztof Hermann	1003\;Hermann Krzysztof	r
21\;przyroda Lubelszczyzna	21\;przyroda I 21\;lubelszczyzna	g
21\;toksykologia I 1003\;zakrzewski	1003\;zakrzewski	i
2019\;zeszyty naukowe akademii rolniczej we wr	2019\;journal of food science	j
21\;ryby ozdobne	21\;ryby akwariowe	t
2019\;inż rolo	2019\;tech. Rol	f
4\;fundusze strukturalne	4\;fundusze z UE	s
1003\;biochemia	4\;biochemia	3

1035\działkowiec	1003\zeszyty problemowe nauk rolniczych	4
4:\Ekonomia	4:\Socjologia społeczna	

Rys. 2. Przykładowy fragment tabeli TWOLOG.

Poniżej wyjaśniono wyodrębnione typy przeformułowań:

- a** – użytkownik zastosował to samo hasło, używając innego indeksu
- b** – (odświeżenie) użytkownik powtórzył swoje pytanie
- u** – użytkownik uszczegółowił zapytanie
- o** – użytkownik uogólnił zapytanie
- d** – użytkownik wyszukiwał tę samą publikację przy użyciu innego indeksu, ale i innego terminu
- p** – użytkownik zastosował różną pisownię tego samego hasła w tym samym indeksie
- 1** – błędna pisownia początkowego hasła
- 2** – błędna pisownia końcowego hasła
- e** – użytkownik zaczął wyszukiwać w tym samym indeksie według nazwiska współautora
- h** – użytkownik zmienia formę gramatyczną wyrazu w tym samym indeksie
- r** – użytkownik stosuje różną kolejność wyrazów w tym samym indeksie
- g** – rozbicie wielowyrazowego hasła na pojedyncze ze spójnikami „i”, „lub”, „nie”.
- i** – użytkownik rezygnuje z wyszukiwania przy użyciu operatorów Boole’a
- j** – inny język – użytkownik stosuje tłumaczenie
- t** – użytkownik stosuje podobne skojarzeniowo hasła
- f** – użytkownik stosuje inne hasło w tym samym indeksie
- s** – użytkownik stosuje synonim w tym samym indeksie
- 3** – użytkownik zastosował błędnie pierwszy indeks (powinien inaczej wyszukiwać)
- 4** – użytkownik zastosował błędnie drugi indeks

Czasem jednak dane przeformułowanie mogło być opisane kilkoma literami, np. „1ui” – pierwsze hasło błędne, użytkownik uszczegółowił pytanie, rezygnując z operatorów Boolowskich w formie jednego hasła. W toku przeprowadzonej analizy autorzy dokonali porównania par zastosowanych terminów wyszukiwawczych przez użytkownika w obrębie danej sesji i na tej podstawie powstał zbiór wynikowy zawierający rekordy z wyodrębnionymi typami przeformułowań.

1.3. Proces łączenia przeformułowań w wątki

Pojawiały się czasem przeformułowania niezwiązane tematycznie ze sobą, takie jak w ostatnim wierszu tabeli na Rys. 2. Takie niezwiązane przeformułowanie oznaczało dla nas zakończenie jednego wątku wyszukiwawczego i rozpoczęcie następnego wątku. Procedura ustalania typu przeformułowań była procedurą ręczną (ręcznie wypełnialiśmy tylko ostatnią kolumnę tabeli TWOLOG – Typ przeformułowania). W oparciu o tak przygotowaną tabelę TWOLOG napisany program komputerowy automatycznie wyodrębnił wątki zestawiał je razem z odpowiadającymi przeformułowaniami w tabeli WLOG, której fragment przedstawiamy na Rys. 3.

Wątek	Przeformułowania	A	B	C
S21\S21\S21\S21\S21\K21\K4\K21\K21 \K21\K21\K21\; K21\K21\K21\K21\K21\	b;o;o;u;t;a;b;b;b;b;b;b;b;b;	1	16	15
S21\S21\S21\S21\S21\S21\S21\S21 \S21\S21\S21\S21\; S21\S21\S21\	b;b;b;b;b;b;b;b;u;b;o;b;b;b;s;	1	15	15
S21\S21\S21\S21\S21\S21\S1003\S21\S21 \S21\S21\S21\; S4\	b;b;b;t;b;4t;3a;b;u;b;o;a;	1	13	11
S21\S21\S21\S21\S21\S21\S21\S21 \S21\S21\S4\S4\; S4\S4\	b;t;b;t;t;t;t;t;t;t;t;t;t;t;	1	14	10
S4\S4\S4\S4\S4\S4\S4\S4\S4\S4 \S4\S4\S4\	u;b;b;o;b;u;t;t;t;t;t;t;t;	1	13	0
S4\S4\S4\S21\S21\S4\S21\S21\S21\S21 \K21\21\	r;u;u;o;u;a;s;b;b;bg;	1	11	8

Rys 3. Fragment typów wątków wyszukiwawczych

W pierwszej kolumnie podano skrótowo interfejsy użytkownika (S-indeks, K-słowa kluczowe, H-słowa z haseł) i użyty indeks wyszukiwawczy, przy czym kolejno zadawane pytania oddzielane są znakiem średnika. W drugiej kolumnie każdej parze pytań z pierwszej kolumny odpowiada opisane przez nas przeformułowanie. W kolumnach A, B i C podano liczbę występowania tego typu wątków, długość wątku i liczbę użytych terminów jhp KABA. Na przykład w przedostatnim wierszu zarejestrowano jeden wątek składający się z 13-krotnego wyszukiwania (przeformułowania) w indeksie tytułowym; podczas tego wątku użytkownik dokonał na terminach wyszukiwawczych następujących operacji: uszczegółowienie, dwukrotne odświeżenie, uogólnienie, odświeżenie, uszczegółowienie, sześciokrotnie zastosował hasła skojarzone.

2. PREZENTACJA WYNIKÓW

2.1. Analiza sesji

Przeanalizujemy teraz (Rys. 4), jak wygląda typowa sesja z katalogiem komputerowym. W czasie tej sesji użytkownik przeciętnie zadaje 1,8 wyszukiwań w indeksie i sprawdza około 2 rekordów bibliograficznych. Natomiast tylko 2 razy na 10 przypadków sesji wyszukuje słowa kluczowe, a 1 na 10 dowiaduje się o swój stan konta. Duży współczynnik zmienności w wyszukiwaniach zaawansowanych i wyszukiwaniach w oparciu o słowa kluczowe oznacza duże zróżnicowanie tych operacji w czasie sesji, co oznacza, że zdarzyły się sesje, gdzie wyszukiwano prawie całkowicie w sposób zaawansowany oraz sesje, w których nie używano tej możliwości wcale. Podobnie osoby wyszukujące według słów w hasłach: albo wyszukiwały prawie tylko według słów, albo nie używały tej metody wyszukiwawczej wcale.

	Przeciętna ilość	Odchylenie standard.	Współczynnik zmienności
Indeks	1,760	3,036	1,725
Rekordy bibliograficzne	2,297	5,818	2,533
Zaawansowane	0,005	0,145	29,199
Słowa w hasłach	0,006	0,101	17,713
Słowa kluczowe	0,233	1,094	4,693
Stan konta	0,105	0,311	2,951
Inne	0,008	0,199	35,368

Rys. 4. Ekran (interfejs) użytkownika typowej sesji

W czasie swojej sesji w indeksie użytkownik przeciętnie zadaje 5 wyszukiwań w indeksie tytułowym oraz 5 w indeksie autorskim na 10 przypadków sesji. Niezmiernie rzadko, gdyż tylko 16 razy na 1000 przypadków sesji, wyszukuje, posiłkując się terminami z języka haseł przedmiotowych. Duży współczynnik zmienności w wyszukiwaniach z użyciem haseł przedmiotowych w językach angielskim i francuskim i wyszukiwaniach w oparciu o sygnaturę oznacza duże zróżnicowanie tych operacji w czasie sesji.

	Przeciętna ilość	Odchylenie standard.	Współczynnik zmienności
Indeks autorski	0,391	1,324	3,384
Sygnatura	0,016	0,307	18,692
Hasło przedm.	0,164	0,924	5,619
Hasło przedm. ang./franc.	0,002	0,055	24,122
Tytuł czasopisma	0,130	1,124	8,642
Indeks tytułowy	0,462	1,734	3,755

Rys. 5. Transakcje użytkownika na ekranie indeks

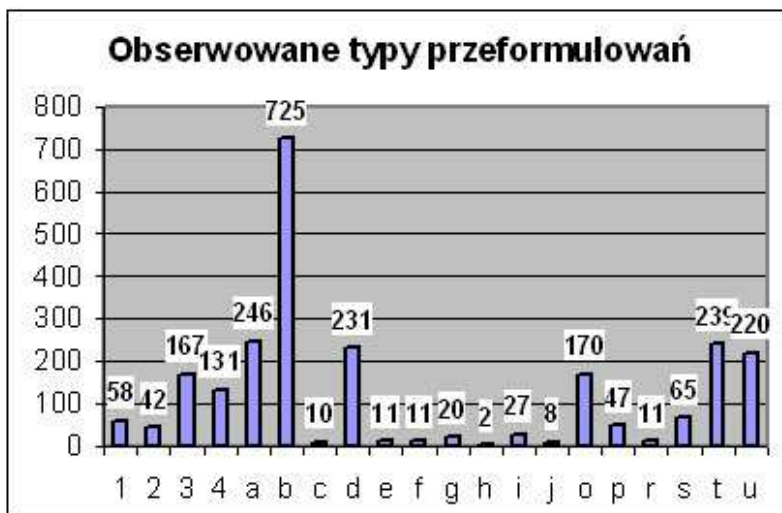
Natomiast przeciętne transakcje użytkownika na ekranie - słowa kluczowe przedstawiamy na Rys. 6.

	Przeciętna ilość	Odchylenie standard.	Współczynnik zmienności
Autor	0,024	0,268	10,985
Wszędzie	0,090	0,581	6,480
Hasło przedmiotowe	0,016	0,255	15,543
Tytuł	0,017	0,159	9,262

Rys. 6. Transakcje użytkownika na ekranie słowa kluczowe

2.2. Analiza przeformułowań

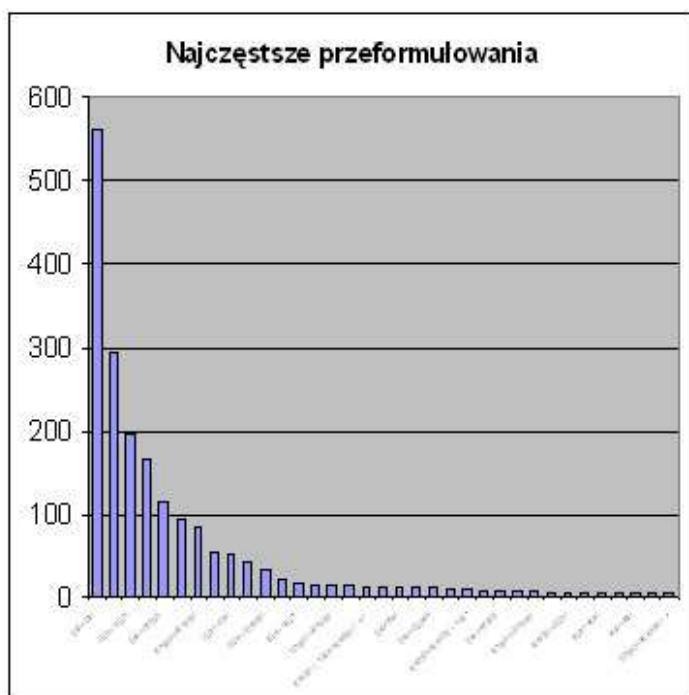
Rys. 7 zestawia rodzaje przeformułowań razem z ich liczbą.



Rys. 7. Obserwowane rodzaje przeformułowań.

Ponieważ niektóre przeformułowania występowały łącznie, dlatego ich sumaryczna liczba jest większa od liczby wszystkich przeformułowań. W kolejności malejącej najczęściej występuje: *odświeżenie* oraz na zdecydowanie niższym, ale wzajemnie zbliżonym, poziomie *identyczne hasło inny indeks*, *hasła skojarzone*, *ta sama pozycja szukana innym indeksem*, *uszczegółowienie* itd. Przeformułowania możemy połączyć w pewne grupy i porównanie tych grup wydaje się ciekawe. Na przykład *uszczegółowienie* (220) jest znacznie częstsze niż *uogólnienie* (170), a więc użytkownicy częściej przechodzą od ogólniejszych pytań do bardziej szczegółowych niż na odwrót. *Łączenie w hasło wielowyrazowe* (27) jest częstsze od operacji rozbicia (20). Częściej *poprawiano błędy* (58) niż *robiono błędy* (42), a ponieważ liczba błędów jest stała, znaczy to, że często w wyszukiwaniu błąd występował na samym początku sesji czy wątku. Częściej najpierw użyto *błędneho indeksu* (167) niż *przeformulowano w błędny indeks* (131), a więc znowu błędnie użyty indeks to często pierwsze zadane pytanie w wątku wyszukiwawczym. W przypadku zmiany indeksu trochę częstsza jest zmiana indeksu z identycznym hasłem (246) niż zmiana indeksu i zmiana hasła (231). Gdy użytkownik pozostawiał ten sam indeks, to aż w 725 przeformułowanie polegało na odświeżeniu ekranu (powtórzenie tego samego hasła), w 230 sytuacjach użyto hasła skojarzonego, w 65 sytuacjach użyto synonimu, w 47 przypadkach była to różna pisownia tego samego hasła, a pozostałe sytuacje były już bardzo rzadkie.

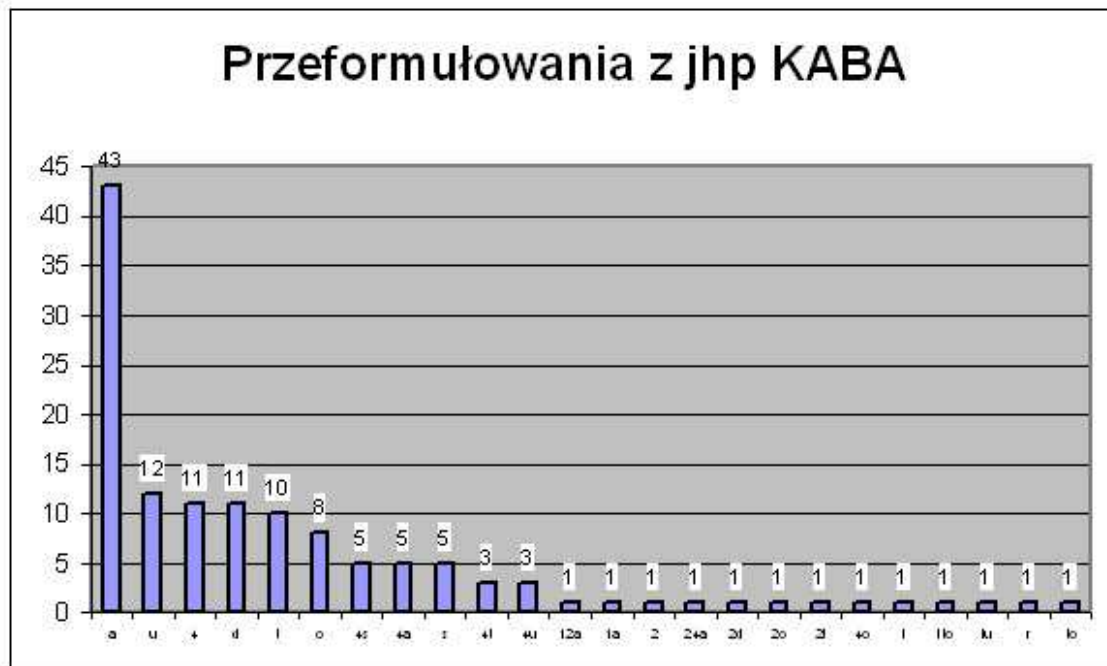
Ponadto przeanalizowaliśmy (Rys. 8.) najczęstsze przeformułowania z punktu widzenia interfejsu i transakcji użytkownika. Wykres pokazuje że najczęściej stosowano kombinacje przejść przeformułowania z użyciem indeksu z tytułu na inny lub ten sam tytuł (561), z użyciem indeksu z autora w autora (295), z użyciem interfejsu indeks z hasła przedmiotowych KABA na te same hasła przedmiotowe KABA (196), z użyciem indeksu z autora w tytuł (164) i vice versa (115). W 95 przypadkach w ramach interfejsu indeks użytkownik przeszedł od tytułu czasopisma do tytułu czasopisma, a w 84 sytuacjach przejście od słów kluczowych wszędzie do słów kluczowych wszędzie, a w 53 i 51 przypadkach przechodzono od tytułu do hasła przedmiotowych i na odwrót.



Rys. 8. Najczęstsze rodzaje przeformułowań.

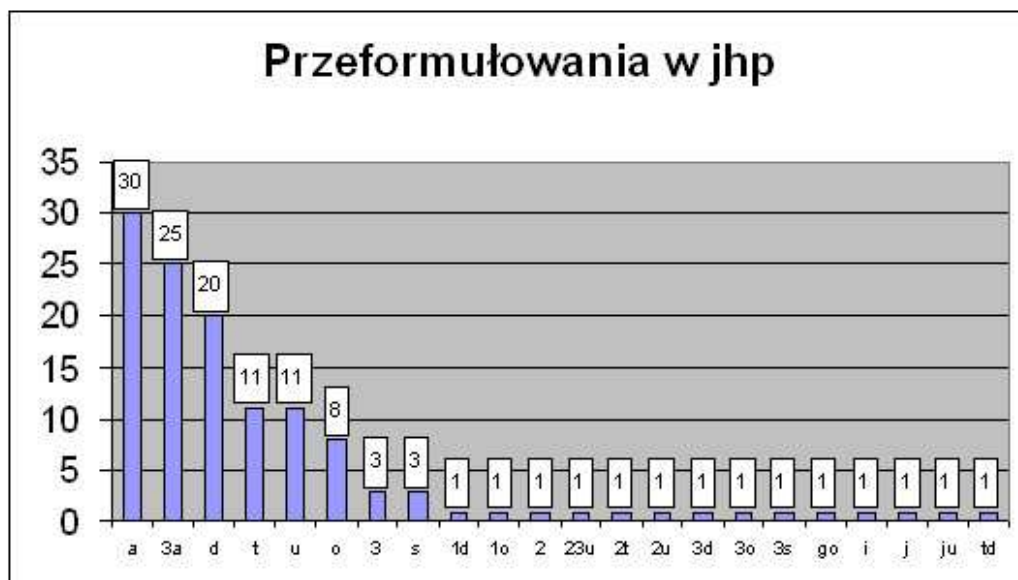
Jako kolejny wynik omówimy przeformułowania z jhp KABA w jakiś inny, dowolny typ

wyszukiwań. Najczęściej użytkownicy z jhp KABA „uciekali” do ekranu indeks - tytuł (51), indeks - autor (35) oraz słowa kluczowe - wszędzie (14). Natomiast sytuacje, w jakich użytkownik sięga do jhp KABA od innych typów wyszukiwań, to w 53 sytuacjach jest przejście od interfejsu indeks - tytuł, a w 44 od indeks - autor. Pozostałe przejścia są już bardzo rzadkie. Pokażemy teraz przeformułowania w ramach jhp KABA. W 196 sytuacjach było to przeformułowanie w ramach ekranu indeks z hasła przedmiotowe w hasła przedmiotowe, a w 17 w ramach ekranu słowa kluczowe z hasła przedmiotowe w hasła przedmiotowe. Pozostałe przeformułowania były sporadyczne. Dalej przeanalizowaliśmy (Rys. 9) rodzaje przeformułowań związanych z jhp KABA.



Rys. 9. Rodzaje przeformułowań z jhp KABA.

Zdecydowanie najczęstszy obserwowany typ przeformułowania z jhp KABA jest to takie same hasło, ale zmiana indeksu. Mogą to być pomyłkowe użycia indeksu (tytuł, autor, zob. rys. 7), ale może też być przeformułowanie jhp KABA w słowa kluczowe wszędzie, a więc niezadowolone z efektów wyszukiwania. Zdecydowanie rzadsze jest uszczegółowienie (12), pomyłka w użyciu następnego indeksu (11), poszukiwanie tej samej pozycji innym indeksem (11), skojarzone hasła (10) i uogólnienie (8).



Rys. 10. Rodzaje przeformułowań w jhp KABA

Z analizy widać, że do jhp KABA przechodzimy, nie zmieniając hasła, a tylko zmieniając indeks (55 przy czym 25 to błędnie użyty pierwszy indeks). Użytkownicy najczęściej podawali hasło przedmiotowe w rubryce tytuł lub autor i dopiero potem przechodzili do haseł przedmiotowych. Duża liczba przeformułowań polega na poszukiwaniu tej samej pozycji innym indeksem (20) i w porównaniu z poprzednimi wynikami widzimy, że najczęściej wyszukiwania według tytułu i autora nie dały dobrych rezultatów. Następną grupą są uszczegółowienia (11) i uogólnienia (8).

Przeformułowania w ramach jhp KABA to w 107 przypadkach odświeżenie ekranu, w 45 hasła skojarzone, w 24 uszczegółowienie, w 23 uogólnienie i w 15 użycie synonimów. W 4 przypadkach używano również różnej pisowni. Hasła skojarzone, synonimy i różna pisownia świadczą o kłopotach użytkowników przy wyszukiwaniu hasłami przedmiotowymi.

2.3. Analiza błędów użytkownika

Przeanalizujemy teraz błędy popełniane przez użytkowników w trakcie wyszukiwania. Zaczniemy od błędnie użytych indeksów.

Błędnie użyty 1 indeks		Liczba	Procent
Autor	Tytuł	82	49,39759
Autor	Hasło przedm	30	18,07229
Autor	Tytuł czasopisma	18	10,84337
Autor	Autor	14	8,433735
Tytuł	Autor	3	1,807229
Tytuł	Tytuł	3	1,807229
Autor	Wszędzie słowa kluczowe	2	1,204819
Słowo kluczowe w haśle autorskim	Słowo kluczowe w tytule	2	1,204819
Słowo kluczowe w tytule	Słowo kluczowe wszędzie	1	0,60241
Autor	Słowo kluczowe wszędzie	1	0,60241
Słowo kluczowe w haśle autorskim	Słowo kluczowe w haśle autorskim	1	0,60241
Złożone	Złożone	1	0,60241
Autor	Słowo kluczowe w tytule	1	0,60241
Sygnatura	Tytuł czasopisma	1	0,60241
Tytuł	Tytuł czasopisma	1	0,60241
Hasło przedm ang./franc.	Hasło przedm	1	0,60241
Tytuł	Hasło przedm	1	0,60241
Tytuł czasopisma	Autor	1	0,60241
Autor	Złożone	1	0,60241
Tytuł	Sygnatura	1	0,60241

Rys. 11. Błędy popełniane przez użytkowników

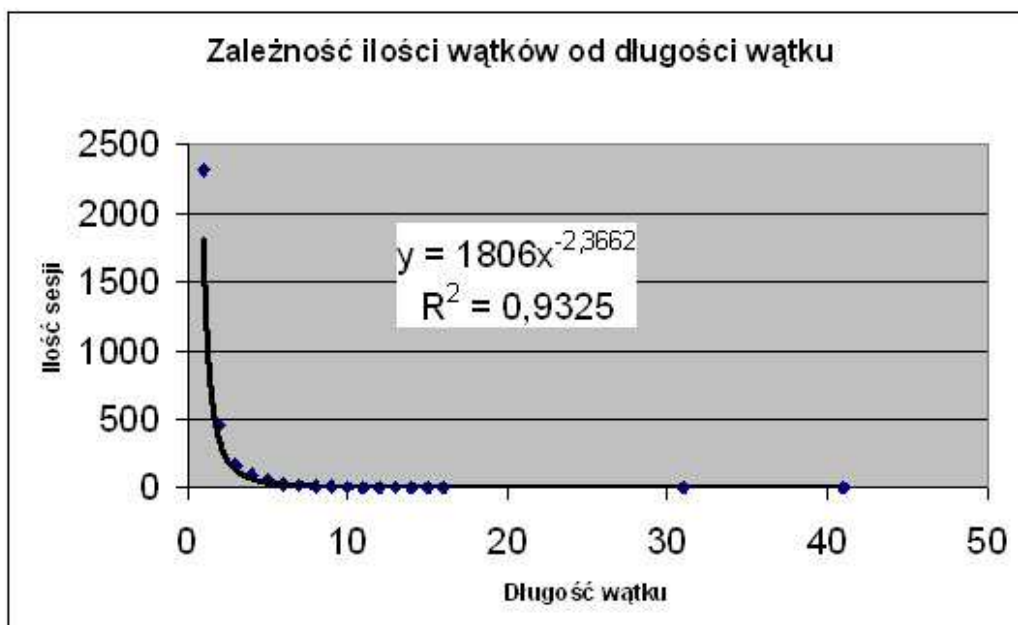
Najczęściej popełnianym błędem jest użycie autora zamiast tytułu (aż 82 sytuacje) oraz autora zamiast haseł przedmiotowych i tytułu czasopisma (łącznie prawie 80% błędów). Oczywiście błędnie użyty indeks autorski związany jest z faktem, że indeks ten proponuje komputer BG AR, jest to indeks domyślny. Może lepiej dla wyszukiwania byłoby, gdyby wszystkie indeksy użytkownik musiał wybierać świadomie. Na uwagę zasługują sytuacje, gdy pierwszym błędnie użytym indeksem jest indeks autorski, a drugim też indeks autorski (lub tytułowy). Oznacza to, że użytkownik dalej popełniał błąd źle użytego indeksu. Analizując, kiedy użytkownik popełnia błąd, odkrywamy, że zazwyczaj jest to ekran indeks – autor (141), ekran indeks – tytuł (8) i ekran słowa kluczowe – autor (3). Ucieczka z błędnie użytego indeksu to w 117 sytuacjach poprawa, w 13 rezygnacja z dalszych wyszukiwań, w 10 zmiana indeksu, ale i zmiana sposobu wyszukiwania, w 6 zmiana indeksu z uogólnieniem, w 4 zmiana indeksu z uszczegółowieniem, a w 4 zmiana indeksu i użycie synonimu.

Zastanówmy się teraz nad pytaniem, kiedy użytkownik popełnia błędy źle użytego indeksu? Okazuje się, że najczęściej przechodząc z ekranów indeks-tytuł (47), indeks-hasła przedmiotowe (27), indeks-autor (21) i indeks-tytuł czasopisma (13). Najczęściej jest to nowy wątek (34), identyczne hasło i zmiana mimowolna indeksu (20), uszczegółowienie (17), synonim (10) i uogólnienie (9).

Przejdźmy teraz do innego rodzaju błędów popełnianych przez użytkownika: prawidłowo użyta metoda wyszukiwawcza, ale błąd w hasło napisanym przez użytkownika. Rozważać tutaj będziemy nie bezwzględną liczbę błędów, ale procentową w stosunku do tego typu wyszukiwań. Względnie najwięcej błędów użytkownicy popełniali w przypadku używania słów kluczowych w tytule (0,07), słowa kluczowe w hasło przedmiotowym (0,05), słowa kluczowe wszędzie (0,02) oraz słowa kluczowe w tytule i indeks autorski (po 0,01). Co ciekawe, błędy te następowały najczęściej wtedy, gdy poprzednim wyszukiwaniem było słowa kluczowe w hasło przedmiotowym (0,02). Najczęściej błędy w pisowni pojawiały się przy zmianie indeksu i na początku wątku wyszukiwawczego.

2.4. Analiza wątków wyszukiwawczych

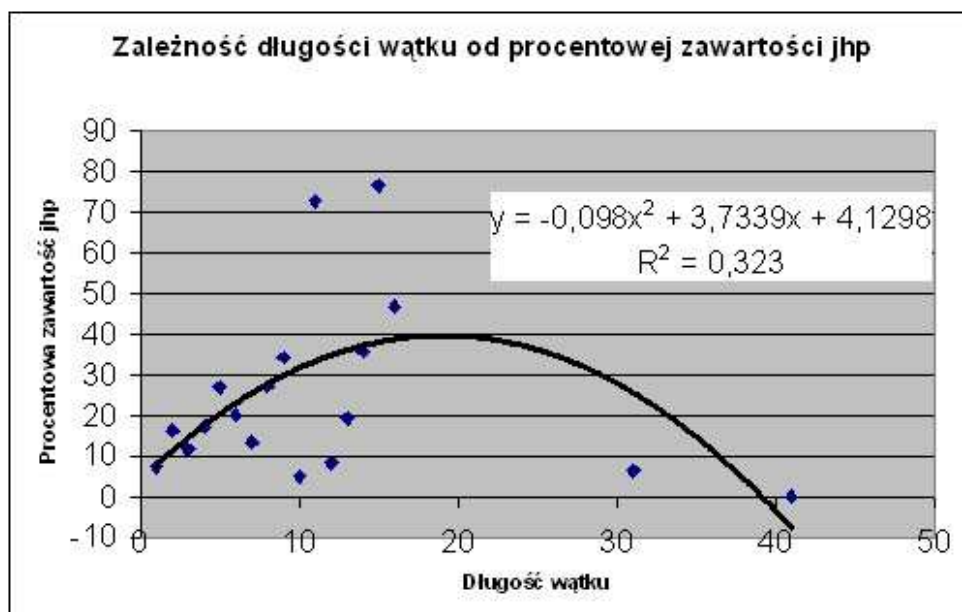
Analizując materiał, jak często pojawiają się wyszukiwania o długich wątkach, zauważyliśmy prawidłowość bardzo zbliżoną do I prawa Zipfa.



Rys. 12. Zależność ilości wątków od długości wątku

Prawidłowość tę prezentujemy na Rys. 12. Zastanawialiśmy się również, czy liczba odwołań do jhp KABA w trakcie wątku nie zależy od długości wątku. Dość nieoczekiwaną zależność prezentujemy na Rys. 13.

Widać, że w miarę wzrostu długości wątku do około 20 rośnie chociaż coraz wolniej liczba użytych pytań z haseł przedmiotowych; potem haseł przedmiotowych jest coraz mniej. Nieoczekiwane jest dość dobre dopasowanie krzywej do danych empirycznych (korelacja 0,323). Widać na wykresie dwa wątki o długości 11 i 15 wypełnione w ponad 70% pytaniami z haseł przedmiotowych. Jednak w tym samym zakresie wątku długości 10 i 12 hasła przedmiotowe występowały mniej niż 10%. Tego typu rozrzut nie występował wcześniej.



Rys. 13. Zależność długości wątku od procentowej ilości użytych w nim wyszukiwań według jhp KABA

3. Wnioski

Analizując otrzymany materiał, obserwujemy wyjątkowo częste błędy w typach wyszukiwań: *słowa kluczowe w hasłach przedmiotowych*. Jest chyba jakaś niezachwiana wiara użytkowników w to, że pisownia haseł przedmiotowych może przypominać korespondencję w Gadu-Gadu. Bardzo często użytkownicy nie zauważają, według jakiego indeksu wyszukują, liczba tego typu błędów jest bardzo duża. Do haseł jhp KABA użytkownicy sięgają, gdy nie udało się dobrze wyszukać według tytułu (rzadziej według autora). Przejście do jhp KABA zbyt często jest efektem błędnie użytego indeksu (pytamy o hasło przedmiotowe w okienku wyszukiwawczym tytuł lub autor). Przejście z jhp KABA do innych typów wyszukiwania również zbyt często jest efektem błędu. Mało jest wątków wyszukiwawczych złożonych w dużym stopniu z pytań jhp KABA; są to przeważnie wątki dłuższe – około 10-15 pytań użytkownika. Jest bardzo niewielka liczba zdecydowanych entuzjastów wyszukiwania wg jhp KABA i ogromna większość osób, które tego wyszukiwania unikają.

Bibliografia:

- [1] BANKS, J. Are Transaction Logs Useful? A Ten-Year Study. *Journal of Southern Academic and Special Librarianship* 2000, 01.
- [2] BURTON, M. C., WALTHER, J. B. The value of Web log data in use-based design and testing. In *Journal of Computer Mediated Communication* [on-line] 2001, 6(1) [dostęp 17 września 2006]. Dostępny w World Wide Web: <http://www.ascusc.org/jcmc/vol6/issue3/burton.html>.
- [3] PETERS, T. A. The history and development of transaction log analysis. *Library Hi Tech* 1993, 42(11), s. 41-66.
- [4] PETERS, T.A. Using Transaction Log Analysis for Library Management Information. *Library Administration and Management* 1996, 10, s. 20-25.
- [5] SLONE, D. J. Encounters with the OPAC: On-line searching in public libraries. *Journal of the American Society for Information Science* 2000, 51 (8), s. 757.
- [6] SOO YOUNG RIEH, A., HONG, (Iris) X. Analysis of multiple query reformulations on the web: the interactive information retrieval context. *Information Processing and Management* 2006, 42, s. 751-768.

[Poprzedni](#) - [Spis treści](#) - [Następny](#)

(C) 2006 EBIB

Schematy zachowań użytkowników języka haseł przedmiotowych KABA / Andrzej Krajka, Artur Ściborek // W: Opracowanie przedmiotowe dokumentów z zakresu nauk ścisłych: matematyczno-przyrodniczych i technicznych. Język haseł przedmiotowych KABA: teoria, praktyka, przyszłość. Kazimierz Dolny, 20-22 września 2006 roku. - [Warszawa] : Stowarzyszenie Bibliotekarzy Polskich, K[omisja] W[ydawnictw] E[lektronicznych], Redakcja "Elektronicznej Biblioteki", 2006. - (EBIB Materiały konferencyjne nr 15). - ISBN 83-921757-6-X. -Tryb dostępu : http://www.ebib.info/publikacje/matkonf/kaba/krajka_sciborek2.php