

JAK CZYTELNIK PORUSZA SIĘ PO BIBLIOTECE CYFROWEJ? ANALIZA WZORCÓW ZACHOWAŃ

Agnieszka Lewandowska

Poznańskie Centrum Superkomputerowo-Sieciowe

Marcin Werla

Poznańskie Centrum Superkomputerowo-Sieciowe



Cel prezentacji

The image displays a terminal window titled "Terminal — ssh — 162x39" with two tabs: "bash" and "ssh". The terminal output consists of a series of log entries, each representing an HTTP request. The logs include the following information for each entry:

- IP address and port (e.g., 65.55.104.14 - -)
- Timestamp (e.g., [09/Dec/2009:00:04:50 +0100])
- HTTP method and version (e.g., GET /style/common/js/jquery-ajax-queue-1.0.js HTTP/1.1)
- Status code (e.g., 200)
- Request size (e.g., 2934)
- Referer (e.g., http://fbc.pionier.net.pl/owoc/results?queryType=-6&query=\Wilno\&isExpandable=off&isRemote=on&dirids=1&roleId=subject&server:type=dlibra&action=DistributedSearchAction)
- User-Agent (e.g., Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4325; .NET CLR 2.0.40607; .NET CLR 3.0.04506.648))

On the right side of the terminal window, a network diagram is overlaid. It features a central node labeled "fbc.pionier.net.pl" which is connected to numerous other nodes. These nodes are arranged in a roughly circular pattern around the center and are labeled with various domain names and IP addresses, such as "www.google.com", "www.msn.com", "www.wikipedia.org", "www.yahoo.com", "www.fox.com", "www.abc.com", "www.cnn.com", "www.foxnews.com", "www.msnbc.com", "www.fox.com", "www.abc.com", "www.cnn.com", "www.foxnews.com", "www.msnbc.com". The connections between nodes are represented by lines, indicating network links or data flow.

Plan prezentacji

1. Metody badań użytkowników
2. Zastosowane podejście
3. Przygotowanie danych
4. Wizualizacja danych
5. Wzorce zachowań
6. Podsumowanie



JAK BADA SIĘ UŻYTKOWNIKÓW?



VI Warsztaty "Biblioteki cyfrowe" 2009

Metody badań

- Bezpośrednie
 - Ankiety (statystyczna)
 - Badania użytkowników w warunkach laboratoryjnych
 - *Hallway testing*
 - *Eye tracking systems*
- Pośrednie
 - Dane z przeglądarki
 - Dane z serwera aplikacji



Badania w praktyce

- Ankieta “*Drogi czytelniku, jak korzystasz z biblioteki cyfrowej?*”
- Statystyki z oprogramowania do tworzenia bibliotek cyfrowych
- Statystyki gromadzone przez darmowe oprogramowanie ogólnego zastosowania
 - ▣ Korzystające z danych przeglądarki
 - ▣ Korzystające z logów serwera aplikacji



Statystyki w dLibrze

Oceny użytkowników

Publikacje grupowe | Wydania

sortuj wg: [ocen pozytywnych](#) | [ocen negatywnych](#)

1. [3] [0] [Heraldyka polska wieków średnich](#)
2. [3] [0] [Powiat pyzdrski w XVI stuleciu](#)
3. [2] [0] [Podstawy Automatyki](#)
4. [2] [0] [Ziółko, Jerzy, 1986, Zbiorniki metalowe na ciecze i gazy](#)
5. [2] [0] [Oświata; pismo sześciotygodniowe poświęcone domowemu i szk](#)
6. [1] [0] [W.A. Mozart, Requiem : für 4 Singstimmen, 2 Violinen, Viola, B; Trompeten, Pauken und Orgel : KV 626 : \[partytura\]](#)
7. [1] [0] [Księga adresowa wszystkich miejscowości w Wielkim Księstwie F](#)
8. [1] [0] [Oleje, paliwa i smary dla motoryzacji i przemysłu](#)
9. [1] [0] [Dwory polskie w Wielkim Księstwie Poznańskim; Polish Manor](#)
10. [1] [0] [Drgania mechaniczne](#)

Najczęściej czytane publikacje : Wielkopolska Biblioteka Cyfrowa

Poniżej znajdują się statystyki biblioteki cyfrowej. Statystyki te zbierane są od dnia: 2004-06-10

Objaśnienia:

W nawiasach kwadratowych przy tytule publikacji podano liczbę wyświetleń danej publikacji.

Wyświetl najczęściej czytanych publikacji.

Strona 1 z 3626

 1 | 2 | 3 | ...

1. [Ocenianie osiągnięć uczniów przez nauczycieli wychowania fizycznego](#) [281688]
2. [Nauczyciel i uczeń wyzwaniem dla przyszłości](#) [118261]
3. [Zasady leczenia i rehabilitacji chorób układu oddechowego w specjalistycznym szpitalu w Ludwikowie](#) [117433]
4. [Digitalizacja zasobów bibliotecznych](#) [77677]
5. [Szermierka w Polsce w latach 1945-1989](#) [71671]
6. [Wybrane psychologiczne aspekty aktywności sportowej. Monografia nr 343.](#) [68784]
7. [Theatrum chemicum, praecipuos selectorum auctorum tractatus de chemiae et lapidis philosophici antiquitate, veritate, iure, praestantia, & operationibus continens. T. 1](#) [64447]
8. [Kultura fizyczna antycznej Grecji. Monografia nr 317.](#) [58344]
9. [Theatrum chemicum, praecipuos selectorum auctorum tractatus de chemiae et lapidis philosophici antiquitate, veritate, iure, praestantia, & operationibus continens. T. 5](#) [57985]
10. [Theatrum chemicum, praecipuos selectorum auctorum tractatus de chemiae et lapidis philosophici antiquitate, veritate, iure, praestantia, & operationibus continens. T. 3](#) [55910]
11. [Biblia święta, tho iest, Księgi Starego y Nowego Zakonu, wlasnie z Zydowskiego, Greckiego, y Łacińskiego, nowo na Polski język z pilnością y wiernie wyłożone \[tzw. "Biblia brzeska"\] \[Stary Testament\]](#) [52656]
12. [Struktura zdolności motorycznych oraz jej globalne i lokalne przejawy](#) [52143]
13. [Theatrum chemicum, praecipuos selectorum auctorum tractatus de chemiae et lapidis philosophici antiquitate, veritate, iure, praestantia, & operationibus continens. T. 4](#) [50137]
14. [Theatri chemici volumen sextum, theologis, medicis, et tam vulgaribus quam hermeticae, chemiae & studiosis](#)

GoogleAnalytics

Google Analytics

fb.c.pc.s.s@gmail.com | Settings | My Account | Help | Sign Out

Analytics Settings | View Reports: fbc.pionier.net.pl

My Analytics Accounts: PSNC DL Team

- Dashboard
- Intelligence Beta
- Visitors
- Traffic Sources
- Content
- Goals

- Custom Reporting Beta

My Customizations

- Custom Reports Beta
- Advanced Segments Beta
- Intelligence Beta
- Email

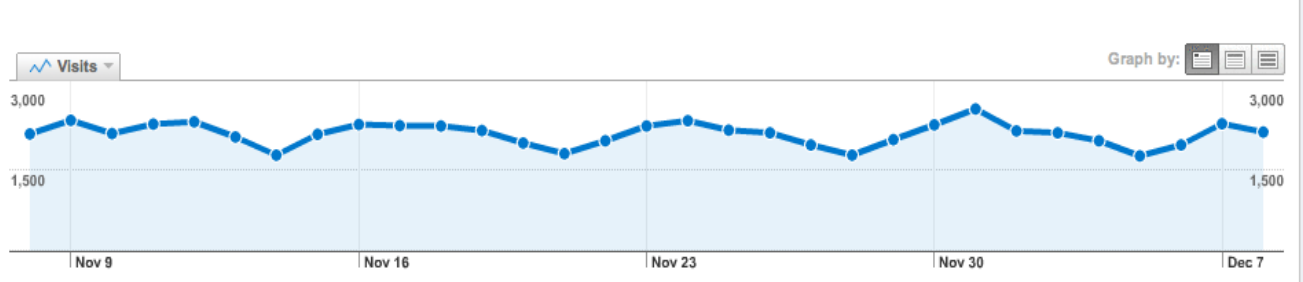
Help Resources

- About this Report
- Conversion University
- Common Questions

Export | Email | Beta Advanced Segments: All Visits

Dashboard

Nov 8, 2009 - Dec 8, 2009



Site Usage

67,172 Visits	40.11% Bounce Rate
386,156 Pageviews	00:06:12 Avg. Time on Site
5.75 Pages/Visit	49.16% % New Visits

Visitors Overview

Map Overlay

AWStats

Statistics for:
www.fbc.pionier.net.pl

[Summary](#)

When:

[Monthly history](#)

[Days of month](#)

[Days of week](#)

[Hours](#)

Who:

[Countries](#)

☐ [Full list](#)

[Hosts](#)

☐ [Full list](#)

☐ [Last visit](#)

☐ [Unresolved IP Address](#)

[Robots/Spiders visitors](#)

☐ [Full list](#)

☐ [Last visit](#)

Navigation:

[Visits duration](#)

[File type](#)

[Viewed](#)

☐ [Full list](#)

☐ [Entry](#)

☐ [Exit](#)

[Operating Systems](#)

☐ [Versions](#)

☐ [Unknown](#)

[Browsers](#)

☐ [Versions](#)

☐ [Unknown](#)

Last Update: 09 Dec 2009 - 17:00

Reported period: Dec 2009 OK

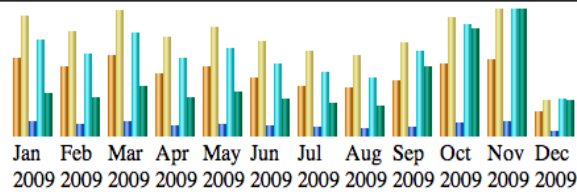


Summary

Reported period	Month Dec 2009				
First visit	01 Dec 2009 - 00:00				
Last visit	09 Dec 2009 - 17:00				
	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Viewed traffic *	14022	20577 (1.46 visits/visitor)	272167 (13.22 Pages/Visit)	2257006 (109.68 Hits/Visit)	21.93 GB (1117.73 KB/Visit)
Not viewed traffic *			945479	950110	21.02 GB

* Not viewed traffic includes traffic generated by robots, worms, or replies with special HTTP status codes.

Monthly history



Month	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Jan 2009	44847	69158	863711	5862124	26.26 GB
Feb 2009	39525	60226	696428	5021284	24.37 GB

ZASTOSOWANE PODEJŚCIE (w teorii)



Co chcemy wiedzieć?

Jak czytelnik korzysta
z biblioteki cyfrowej?

Jakie strony przegląda?
Skąd do nich przyszedł?
Gdzie z nich odchodzi?



Założenia

- Brak ingerencji w oprogramowanie
- Brak ingerencji w zachowania czytelnika
- Jak największa grupa badanych
 - ▣ Idealnie **cała populacja**
- **Tylko** strony związane z funkcjonalnością biblioteki cyfrowej
- Chcemy wiedzieć “**gdzie?**” zamiast “co dokładnie?”



Zastosowane podejście

- Jakie dane?
 - ▣ Logi dostępu serwera aplikacji
- Czyje dane?
 - ▣ Wielkopolska Biblioteka Cyfrowa
- Co dodano do danych?
Wiedzę (!)
 - ▣ o istotnych stronach
 - ▣ o strukturze serwisu



PRZYGOTOWANIE DANYCH



VI Warsztaty "Biblioteki cyfrowe" 2009

Logi dostępu

- Data i czas z dokładnością do sekundy
- Adres IP użytkownika
- Wydane żądanie HTTP
- Status odpowiedzi HTTP
- Ilość przesłanych bajtów w odpowiedzi
- Adres strony (URL) z jakiej wykonano zapytanie (jeżeli istnieje)
- Narzędzie, przy pomocy którego wykonano zapytanie



Logi dostępu

```
Terminal — ssh — 162x39
bash ssh
65.55.104.14 - - [09/Dec/2009:00:04:50 +0100] "GET /style/common/js/jquery-ajax-queue-1.0.js HTTP/1.1" 200 2934 "http://fbc.pionier.net.pl/owoc/results?queryType=-6&query=\Wilno\&isExpandable=off&isRemote=on&dirids=1&roleId=subject&server:type=dlibra&action=DistributedSearchAction" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4325; .NET CLR 2.0.40607; .NET CLR 3.0.04506.648)"
65.55.106.133 - - [09/Dec/2009:00:04:50 +0100] "GET /owoc/results?isExpandable=off&dirids=1&queryType=-6&query=%22Mitteilungen+des+Copernicus-Vereins+f%C3%BCr+Wissenschaft+und+Kunst+zu+Thorn.+H.+45.%22&roleId=title&isRemote=on&server:type=dlibra&action=DistributedSearchAction HTTP/1.1" 200 42482 "-" "msnbot/2.0b (+http://search.msn.com/msnbot.htm)"
66.249.65.148 - - [09/Dec/2009:00:04:50 +0100] "GET /owoc/results?queryType=-6&query=%22Bojownicy%20kap%C5%82ani%20za%20spraw%C4%99%20Ko%C5%9Bcio%C5%82a%20i%20ojczyzny%20w%20latach%201861-1915%20;%20materja%C5%82y%20z%20urz%C4%99dowych%20C5%9Bwiadectw%20w%C5%82adz%20rosyjskich,%20archiw%C3%B3w%20konsystorskich,%20zakonnyc h%20i%20prywatnych.%20Cz.%201,%20Dawne%20Kr%C3%B3l.%20Polskie.%20T.%202,%20diecezje:%20lubelska%20z%20podlask%C4%85%20i%20p%C5%82ocka.%22&isExpandable=off&isRemote=on&dirids=1&roleId=title&hostsId=n43&server:type=dlibra&action=DistributedSearchAction HTTP/1.1" 200 23994 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
65.55.104.14 - - [09/Dec/2009:00:04:51 +0100] "GET /style/common/js/jquery-biframe-2.1.1.js HTTP/1.1" 200 1402 "http://fbc.pionier.net.pl/owoc/results?queryType=-6&query=\Wilno\&isExpandable=off&isRemote=on&dirids=1&roleId=subject&server:type=dlibra&action=DistributedSearchAction" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4325; .NET CLR 2.0.40607; .NET CLR 3.0.04506.648)"
66.249.65.148 - - [09/Dec/2009:00:04:51 +0100] "GET /owoc/results?queryType=-6&query=%22Kresy%20wschodnie%20Rzeczypospolitej%20-%20przemys%C5%82%20drzewny%22&isExpandable=off&isRemote=on&dirids=1&roleId=subject&hostsId=n8&server:type=dlibra&action=DistributedSearchAction HTTP/1.1" 200 22725 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
79.185.61.204 - - [09/Dec/2009:00:04:51 +0100] "GET /owoc/rstats?url=http%3A%2F%2Fdelibra.bg.pol.pl%2F%2F%2Fdocmetadata%3Fid%3Ddoi%3Adelibra.bg.pol.pl%3A29,9%26amp%3Bfrom%3DFBC HTTP/1.1" 302 - "http://fbc.pionier.net.pl/owoc/results?queryType=-2&query=ruby&isExpandable=off&isRemote=on&dirids=1&hostsId=n25&server:type=dlibra&action=DistributedSearchAction" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.5) Gecko/20091102 Firefox/3.5.5"
79.185.61.204 - - [09/Dec/2009:00:04:51 +0100] "GET /owoc/results?queryType=-2&query=ruby&isExpandable=off&isRemote=on&dirids=1&hostsId=n25&server:type=dlibra&action=DistributedSearchAction HTTP/1.1" 200 24724 "http://fbc.pionier.net.pl/owoc/results?queryType=-2&query=ruby&isExpandable=off&isRemote=on&dirids=1&hostsId=n25&server:type=dlibra&action=DistributedSearchAction" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.5) Gecko/20091102 Firefox/3.5.5"
65.55.104.14 - - [09/Dec/2009:00:04:52 +0100] "GET /style/common/js/jquery-tooltip/jquery-tooltip-1.3.css HTTP/1.1" 200 179 "http://fbc.pionier.net.pl/owoc/results?queryType=-6&query=\Wilno\&isExpandable=off&isRemote=on&dirids=1&roleId=subject&server:type=dlibra&action=DistributedSearchAction" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4325; .NET CLR 2.0.40607; .NET CLR 3.0.04506.648)"
65.55.104.14 - - [09/Dec/2009:00:04:52 +0100] "GET /style/common/js/jquery-tooltip/jquery-tooltip-1.3p.js HTTP/1.1" 200 8951 "http://fbc.pionier.net.pl/owoc/results?queryType=-6&query=\Wilno\&isExpandable=off&isRemote=on&dirids=1&roleId=subject&server:type=dlibra&action=DistributedSearchAction" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4325; .NET CLR 2.0.40607; .NET CLR 3.0.04506.648)"
66.249.65.148 - - [09/Dec/2009:00:04:52 +0100] "GET /owoc/results?isExpandable=off&dirids=1&queryType=-6&query=%22171500+II%22&roleId=source&isRemote=on&server:type=dlibra&action=DistributedSearchAction HTTP/1.1" 200 39189 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.65.148 - - [09/Dec/2009:00:04:53 +0100] "GET /owoc/results?queryType=-6&query=%22ilo%C5%9B%C4%87%20skar%C3%B3w;%20241%20;%20sygn.%20%C5%81%20125960%22&isExpandable=off&isRemote=on&dirids=1&roleId=description&hostsId=n44&server:type=dlibra&action=DistributedSearchAction HTTP/1.1" 200 22661 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.65.148 - - [09/Dec/2009:00:04:53 +0100] "GET /owoc/results?queryType=-6&query=%22ilo%C5%9B%C4%87%20skar%C3%B3w;%20241%20;%20sygn.%20%C5%81%20125960%22&isExpandable=off&isRemote=on&dirids=1&roleId=description&hostsId=n18&server:type=dlibra&action=DistributedSearchAction HTTP/1.1" 200 22670 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
PBC09.Gradn
agnieszka lewandowska
```


Przetworzenie danych

- Usunięcie żądań wykonanych przez maszyny
boty, harwestery, crawlery
- Usunięcie żądań o obiekty związane z wyglądem/działaniem strony
pliki CSS, pliki JS
- Usunięcie niepoprawnych żądań
- Identyfikacja poprzedniego żądania
- Przypisanie adresu żądania (URLa) do konkretnej funkcji biblioteki cyfrowej

Dane końcowe

- Żądania HTTP wykonane tylko przez czytelników
- Wyodrębnione sesje czytelników
- Wyodrębnione powiązania między żądaniami
 - skąd → gdzie → dokąd
- Informacja o użytej funkcji biblioteki cyfrowej
 - Istotne przy agregacji danych (!)



WIZUALIZACJA DANYCH



VI Warsztaty "Biblioteki cyfrowe" 2009

- Overview
- Top Content**
- Content by Title
- Content Drilldown
- Top Landing Pages
- Top Exit Pages
- Site Overlay
- Site Search
- Goals

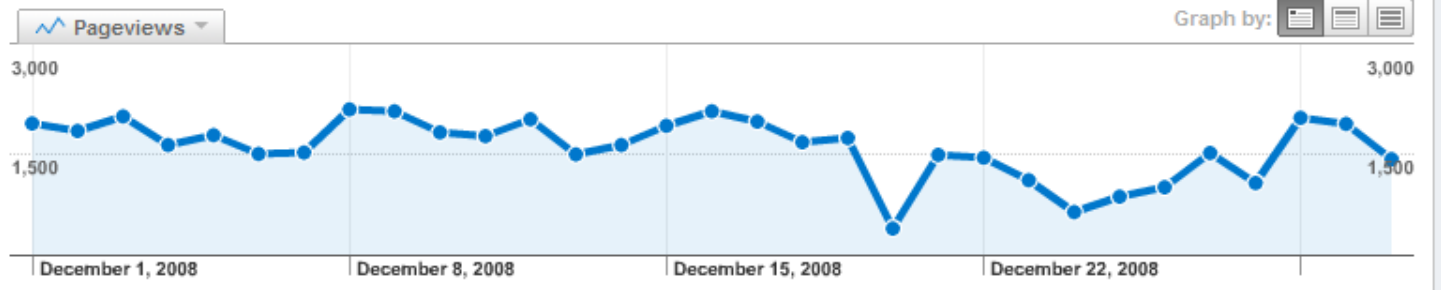
Custom Reporting **Beta**

Settings

- Advanced Segments **Beta**
- Email

Help Resources

- About this Report
- Conversion University
- Common Questions



This page was viewed 49,895 times

Visit this page Analyze: Entrance Paths Content /dlibra

Started here

This Page

Then viewed these pages:


Content	Visits	% visits
/dlibra	1,970	12.46%
/dlibra/latest?dirids=1	1,705	10.78%
/dlibra/text?id=journals	1,020	6.45%
/dlibra/text?id=wojsko	406	2.57%
/dlibra/text?id=addressbucl	314	1.99%
/publication/81942	245	1.55%
/dlibra/collectiondescriptor	240	1.52%
/dlibra/indexsearch?attId=ti	225	1.42%
/dlibra/indexsearch?attId=c	182	1.15%
/dlibra/collectiondescriptor	162	1.02%

Show rows: 10 1 - 10 of 7,005

And ended up here:

Content	Visits	% visits
/dlibra	106	10.41%
/dlibra/text?id=journals	84	8.25%
/dlibra/info?mimetype=image/x.	19	1.87%
/publication/16370	11	1.08%
/publication/2290	9	0.88%
/dlibra/collectiondescription	8	0.79%
/dlibra/info?mimetype=image/x.	5	0.49%
/dlibra/content-sidebar?id=397	3	0.29%
/dlibra/info?handler=djvu&mim	3	0.29%
/publication/publication?id=343	3	0.29%

Show rows: 10 1 - 10 of 749



WIELKOPOLSKA BIBLIOTEKA CYFROWA

readers on-line: 180 | 3.1% | 0.9% DNS | 0% JUNT | 0% | CONTACT

- ### Library
- ▼ Digital Library of Wielkopolska (0%)
 - ▶ Digital heritage (0.9%)
 - > Digital materials (0.3%)
 - > Digital books (0%)
 - > Digital materials (1.6%)
 - Kolekcje wydzielone:
 - > (0%)
 - > The Unity of the Brethren (0%)
 - Archiwalia miast
 - w (0%)
 - > (0.3%) and newspapers
 - > Jordaniana (Henryk Jordan)
 - Jednostki (0.3%)
 - > (0.9%)
 - > Rozprawy doktorskie i
 - h (0.6%)
 - > (4.4%)
 - > Powstanie styczniowe w zbiorach Wielkopolskiej
 - Bi (0%)
 - > (0%)

Digital Library of Wielkopolska | Digital libraries federation

Search: only in text | 3.8% description | in text and in description

Advanced

Proper query syntax

Exhibitions

- Powstanie Wielkopolskie 1918/19
 - Polonia Maior-fontes
-

Recently added

1. Międzynarodowa konkurencyjność polskiej branży ałych
2. Uwarunkowania stabilizowania systemu go przez bank centralny
3. Małe przedsiębiorstwa w Polsce na tle przeobrażeń ch (1944-1991)
4. Kometateza winylotri(metylo,alkoksy)silanów z alkenami katalizowania kompleksami rutenu - Nowa

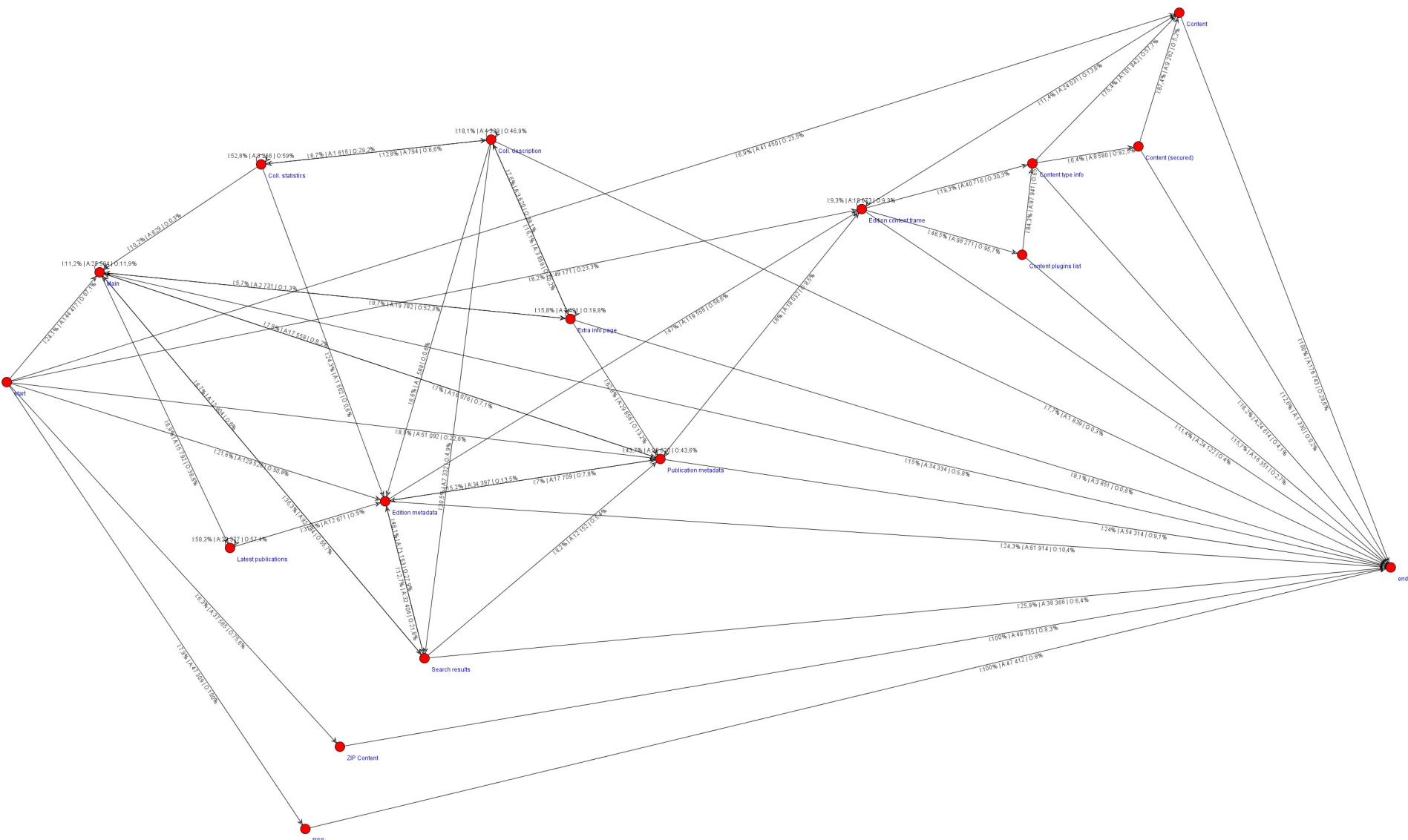
Frequently read publications

1. Ocenianie osiągnięć uczniów przez nauczycieli ia fizycznego [189326]
2. i uczeń wyzwaniem dla przyszłości [103875]
3. Zasady leczenia i rehabilitacji chorób układu oddechowego w specjalistycznym szpitalu w [97296]
4. ja zasobów bibliotecznych [66400]

Wizualizacja danych

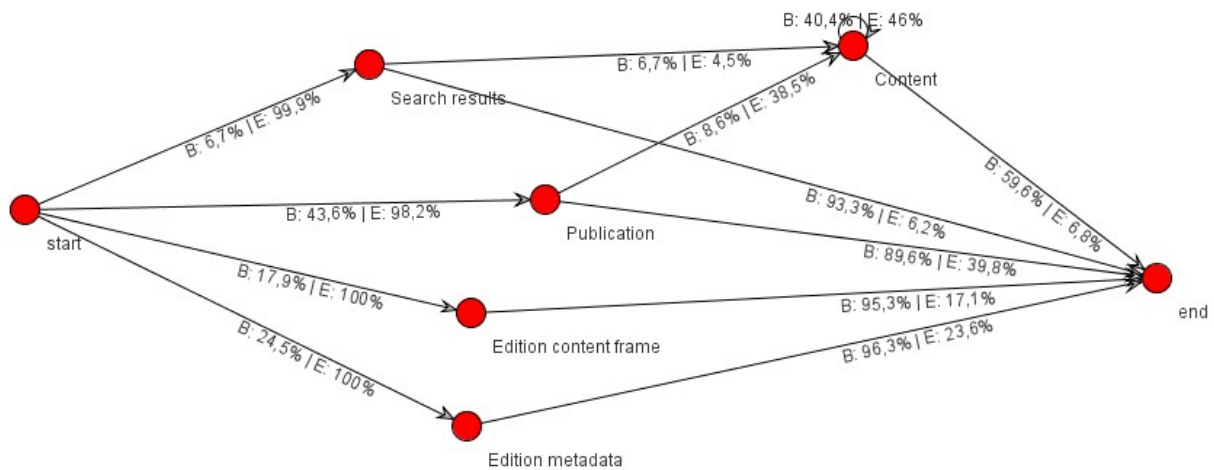
- Dane uzyskane w efekcie obróbki logów dostępu WWW zostały zwizualizowane przy pomocy grafu
 - Wierzchołki – elementy interfejsu czytelnika biblioteki cyfrowej
 - Łuki (łączy wierzchołki) – przejścia czytelników pomiędzy elementami interfejsu biblioteki cyfrowej
 - Opis łuków – % ruchu wchodzącego/wychodzącego przez dany łuk



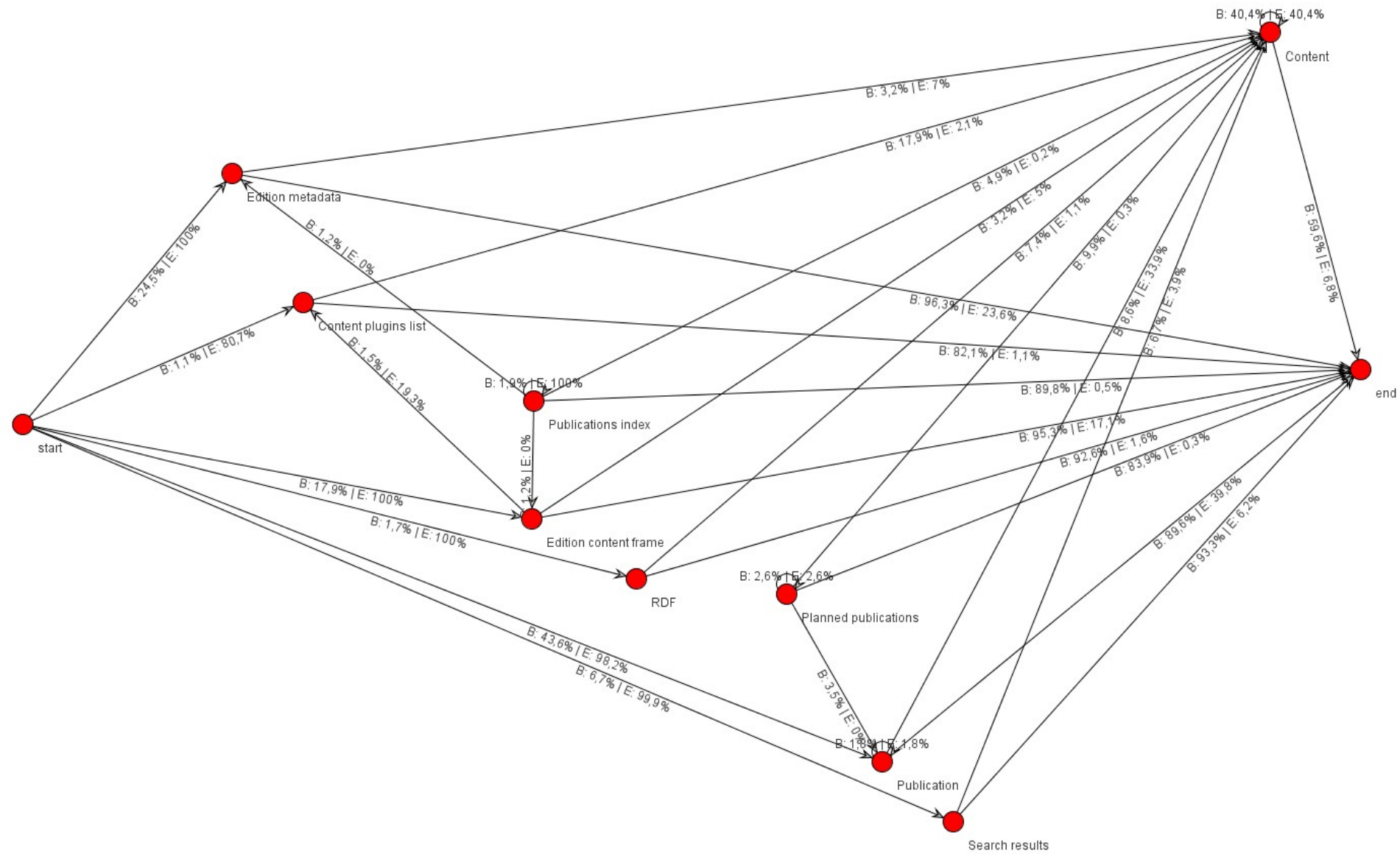


Ludzie, grudzień 2008, min 5% ruchu wychodzącego





Roboty, grudzień 2008, min 5% ruchu wychodzącego

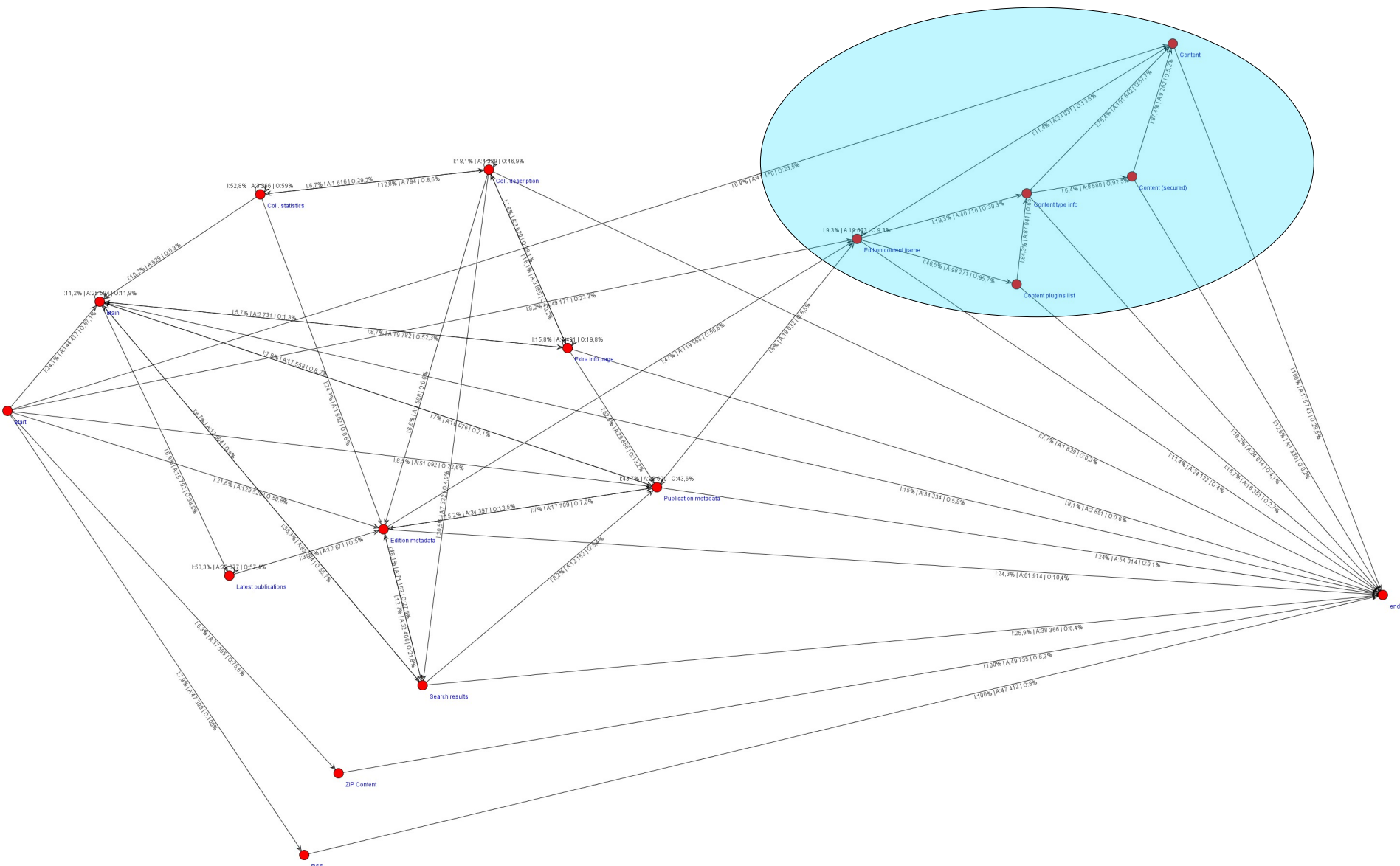


Roboty, grudzień 2008, min 1% ruchu wychodzącego

WZORCE ZACHOWAŃ



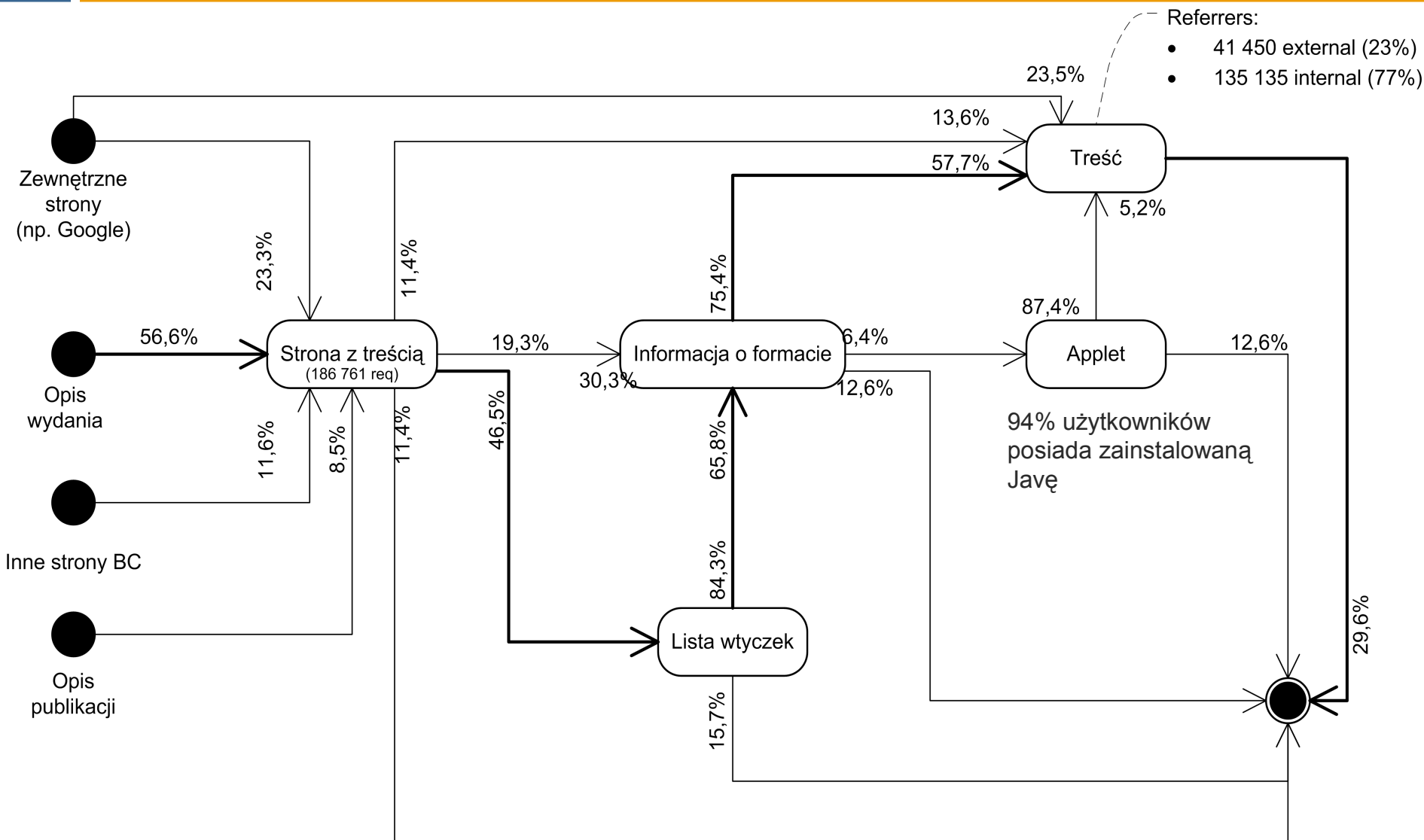
VI Warsztaty "Biblioteki cyfrowe" 2009



Ludzie, grudzień 2008, min 5% ruchu wychodzącego



Przykład analizy wzorców zachowań



DZIĘKUJEMY ZA UWAGĘ! PYTANIA?

Agnieszka Lewandowska

jagna@man.poznan.pl

Marcin Werla

mwerla@man.poznan.pl

