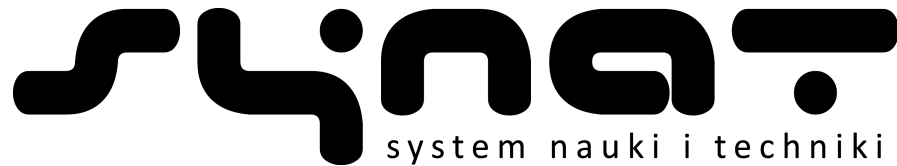


Raport z uruchomienia bazy wiedzy UJ (Jagiellońskiej Biblioteki
Cyfrowej) i jej integracji z systemami zewnętrznymi.



Etap B25

Biblioteka Jagiellońska

Uniwersytet Jagielloński

Kraków 2013

Autor: Bartłomiej Brawuski, Łukasz Mesek, Adrian Wiecheć

Niniejsza praca jest finansowana przez Narodowe Centrum Badań i Rozwoju nr umowy SP/I/1/77065/10 w ramach programu strategicznego: Interdyscyplinarny system interaktywnej informacji naukowej i naukowo technicznej.

Zadanie Badawcze SYNAT pt.: „Utworzenie uniwersalnej, otwartej, repozytoryjnej platformy hostingowej i komunikacyjnej dla sieciowych zasobów wiedzy dla nauki, edukacji i otwartego społeczeństwa wiedzy”

Spis treści

1. Cel opracowania
2. Zarządzanie systemem dLibra Jagiellońskiej Biblioteki Cyfrowej
 - 2.1. System monitorowania aplikacji dLibra Jagiellońskiej Biblioteki Cyfrowej
 - 2.2. Dostosowywanie konfiguracji systemu dLibra Jagiellońskiej Biblioteki Cyfrowej rozdział serwisów dLibry na oddzielne serwery
3. Protokół OAI-PMH
 - 3.1. Zapytania protokołu OAI-PMH
 - 3.2. Odpowiedzi protokołu OAI-PMH
 - 3.3. Format metadanych Dublin Core
 - 3.4. Pobieranie metadanych z systemu dLibra za pomocą protokołu OAI-PMH
4. Integracja aplikacji dLibra Jagiellońskiej Biblioteki Cyfrowej z Portalem UJ-BJ
5. Wnioski

1. Cel opracowania

Opracowanie miało na celu przedstawienie sposobu zarządzania bazą wiedzy od strony informatycznej, sposobami eksportu i importu danych, w tym szczególnie metadanych oraz określenie stopnia zainteresowania użytkowników Jagiellońskiej Biblioteki Cyfrowej w odniesieniu do poszczególnych kategorii zasobów informacyjnych, porównanie otrzymanych wyników pochodzących z logów z systemu dLibra z wynikami ankiety przeprowadzonej wśród użytkowników JBC. Poniższe opracowanie jest kolejnym z cyklu opracowań mających określić sposoby integracji Jagiellońskiej Biblioteki Cyfrowej, pełniącej funkcję bazy wiedzy UJ, jak również jednego z repozytorium dziedzinowych w projekcie SyNaT, z platformą PASSIM realizowaną przez NASK w tymże projekcie, ale również z innymi systemami służącymi do gromadzenia i udostępniania dokumentów pełnotekstowych lub też tylko informacji bibliograficznej o nich.

W pierwszej kolejności przedstawiono zarządzanie oprogramowaniem dLibra jako platformą, na której baza wiedzy UJ została posadowiona.

Następnie opisany został protokół OAI-PMH, który jest podstawą wymiany metadanych dla dokumentów gromadzonych w różnego typu repozytoriach, działających na różnego typu oprogramowaniach.

Na koniec przedstawiono sposób integracji bazy wiedzy UJ z portalem Biblioteki Jagiellońskiej oraz portalem głównym uniwersytetu. Został tu opisany sposób na tymczasowe rozwiązanie, do momentu kiedy powstanie lub zostanie zakupione oprogramowanie pozwalające na przeszukiwanie całej domeny uj.edu.pl z jednego miejsca.

2. Zarządzanie systemem dLibra Jagiellońskiej Biblioteki Cyfrowej

Oprogramowanie dLibra (Digital Library Framework) to pierwsze polskie środowisko służące budowie bibliotek cyfrowych rozwijane jest przez Poznańskie Centrum Superkomputerowo-Sieciowe od 1999 roku. Zostało ono stworzone jako środowisko biblioteki cyfrowej nowej generacji, umożliwiającego zarządzanie, przetwarzanie, przeszukiwanie i dostarczanie dokumentów w formie elektronicznej [Wikipedia, 2013]. Z punktu widzenia administratora oprogramowanie to jest bardzo trudno skalowalne, ponieważ dostęp do kodu źródłowego jest zamknięty. W przypadku biblioteki cyfrowej tak dużej jak Jagiellońska Biblioteka Cyfrowa, spełniającej nie tylko funkcje prezentacyjne kopii cyfrowych dokumentów papierowych, ale rozwijającej nowe usługi, oferującej nowe typy dokumentów oraz pełniącej funkcję bazy wiedzy dla uniwersytetu, zamknięty kod źródłowy stanowi problem. Jest to problem zarówno w codziennym zarządzaniu oprogramowaniem jak i wprowadzaniem udogodnień, których oczekują użytkownicy oraz bibliotekarze zarządzający powiększającym się zasobem dokumentów.

2.1. System monitorowania aplikacji dLibra Jagiellońskiej Biblioteki Cyfrowej

Aby określić obciążenie generowane przez oprogramowanie dLibra dla różnych kontekstów czasowych (np. obciążenie godzinne, dzienne, tygodniowe, miesięczne), konieczne było wdrożenie systemu monitoringu, gromadzącego dane na temat wybranych parametrów systemu operacyjnego, obrazujących poziom wykorzystania dostępnych zasobów sprzętowych serwera (np. procentowe wykorzystanie CPU, RAM, I/O urządzeń dyskowych, aktywność interfejsów sieciowych itp.).

W tym celu należało wybrać:

- źródła danych obrazujących obciążenie systemu
- sposób gromadzenia i przechowywania tych danych
- sposób prezentacji zgromadzonych danych

Ponieważ system dLibra Jagiellońskiej Biblioteki Cyfrowej został zainstalowany na platformie Linux, jako narzędzie do monitorowania obciążenia systemu wybrano "sar" z pakietu "sysstat". Po instalacji i konfiguracji pakietu, sysstat gromadzi statystyki systemowe domyślnie w przedziałach czasowych co 10 minut. Pliki z danymi statystyk dziennych są rotowane codziennie o północy.

Monitorowaniu poddano następujące zasoby systemowe:

- procesory (%CPU)
- pamięć (%RAM, swap)
- średnie obciążenie (Load Average)
- liczba otwartych gniazd sieciowych TCP (TCP sockets)
- interfejsy sieciowe (pakiety wysłane kB/s, pakiety odebrane kB/s)
- urządzenia dyskowe (liczba odczytów /s, liczba zapisów/s, %I/O)

Jako narzędzie do graficznej prezentacji danych ze statystyk systemowych (zgromadzonych przy pomocy narzędzia z pakietu sysstat) został wybrany linuxowy pakiet RRDtool. Jest to pakiet który zapisuje dane w seriach czasowych, oraz pozwala wygenerować graficzną reprezentację zgromadzonych serii danych w postaci obrazka (np. jpg lub png).

Poniżej w skrócie opisano przykładową konfigurację systemu monitoringu użycia zasobów dla serwerów obsługujących Jagiellońską Bibliotekę Cyfrową.

Konfiguracja systemu monitoringu obciążenia serwerów

Pierwszym krokiem konfiguracji systemu monitoringu jest wygenerowanie pliku .rrd do gromadzenia próbek wartości parametrów obciążenia serwera, które mają być przedstawione graficznie. W pliku tym zostają zdefiniowane i zainicjowane źródła danych (DS), dla każdego z analizowanych parametrów obciążenia. Określony jest też interwał próbkowania, czyli odstępy czasowe, w jakich będą rejestrowane próbki wartości parametrów.

Poniższy skrypt tworzy plik rrd_db.rrd ,w którym próbki gromadzone będą w odstępach co 10 minut, i będzie przechowywane 6048 próbek, czyli będzie to odpowiadało okresowi 60 dni.

```
#!/bin/bash
start_time=`date -d "04/10/2011 00:00:00" +%s`

# probki co 10 min (-s 600)
#RRA:AVERAGE:0.5:1008 – średnia z jednej próbki; będzie przechowywanych 6048 próbek (60 dni)
# cpu: %user %nice %system %iowait %steal %idle
# mem: kbmempfree kbmempused %memused kbbuffers kbcached kbcommit %commit
# io: tps rd_sec/s wr_sec/s avgrq-sz avgqu-sz await svctm %util
rrdtool create ./rrd_db.rrd -s 600 --start ${start_time} \
DS:c_user:GAUGE:650:0:100 \
DS:c_nice:GAUGE:650:0:100 \
```



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

```
DS:c_system:GAUGE:650:0:100 \  
DS:c_iowait:GAUGE:650:0:100 \  
DS:c_steal:GAUGE:650:0:100 \  
DS:c_total:GAUGE:650:0:100 \  
DS:m_memused:GAUGE:650:0:100 \  
DS:m_commit:GAUGE:650:0:100 \  
DS:i_rd_sec:GAUGE:650:0:U \  
DS:i_wr_sec:GAUGE:650:0:U \  
DS:i_await:GAUGE:650:0:U \  
DS:i_util:GAUGE:650:0:U \  
DS:i0_rd_sec:GAUGE:650:0:U \  
DS:i0_wr_sec:GAUGE:650:0:U \  
DS:i0_await:GAUGE:650:0:U \  
DS:i0_util:GAUGE:650:0:U \  
DS:c0d0_rd_sec:GAUGE:650:0:U \  
DS:c0d0_wr_sec:GAUGE:650:0:U \  
DS:c0d0_await:GAUGE:650:0:U \  
DS:c0d0_util:GAUGE:650:0:U \  
DS:ldavg1:GAUGE:650:0:U \  
DS:ldavg5:GAUGE:650:0:U \  
DS:ldavg15:GAUGE:650:0:U \  
DS:ld_run:GAUGE:650:0:U \  
DS:ld_plst:GAUGE:650:0:U \  
DS:kbswpused:GAUGE:650:0:U \  
DS:eth0_rxkB:GAUGE:650:0:U \  
DS:eth0_txkB:GAUGE:650:0:U \  
DS:lo_rxkB:GAUGE:650:0:U \  
DS:lo_txkB:GAUGE:650:0:U \  
DS:tpsck:GAUGE:650:0:U \  
DS:tcp_tw:GAUGE:650:0:U \  
RRA:AVERAGE:0.5:1:6048
```

Aktualizacja danych w pliku bazy rrd "rrd_db.rrd" jest realizowana przy pomocy skryptu "rrd_update.sh", wywoływanego cyklicznie jako zadanie cron'a, co 10 minut. Zawartość skryptu "rrd_update.sh" przedstawia poniższy listing.

```
#!/bin/bash  
  
rrd_db=rrd_db.rrd  
ftmp_cpu=/tmp/sar_cpu.tmp  
ftmp_mem=/tmp/sar_mem.tmp
```



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

```
ftmp_io=/tmp/sar_io.tmp  
ftmp_io0=/tmp/sar_io0.tmp  
ftmp_c0d0=/tmp/sar_c0d0.tmp  
ftmp_ldavg=/tmp/sar_ldavg.tmp  
ftmp_swp=/tmp/sar_swp.tmp  
ftmp_lo=/tmp/sar_net_lo.tmp  
ftmp_eth0=/tmp/sar_net_eth0.tmp  
ftmp_sock=/tmp/sar_net_sock.tmp
```

```
file=${1}  
t1=`expr index "$file" '/'`  
if [ $t1 -lt 1 ]; then  
  file="/var/log/sysstat/"${file}  
fi
```

```
cpu_cmd="sar -f ${file}"  
mem_cmd="sar -r -f ${file}"  
io_cmd="sar -d -p -f ${file}"  
ldavg_cmd="sar -q -f ${file}"  
swp_cmd="sar -S -f ${file}"  
eth_cmd="sar -n DEV -f ${file}"  
sock_cmd="sar -n SOCK -f ${file}"
```

```
if [ -e ${file} ]; then
```

```
  date1=`${cpu_cmd} | head -1`  
  t1=`expr index "$date1" '/'`  
  if [ $t1 -gt 0 ]; then  
    date1=${date1:$t1-3:10}  
  fi
```

```
  ${cpu_cmd} | grep ':' | grep -v 'CPU' | grep -v 'Average' > $ftmp_cpu  
  ${mem_cmd} | grep ':' | grep -v 'kmem' | grep -v 'Average' > $ftmp_mem  
  ${io_cmd} | grep 'dev254-1' | grep -v 'Average' > $ftmp_io  
  ${io_cmd} | grep 'dev254-0' | grep -v 'Average' > $ftmp_io0  
  ${io_cmd} | grep 'c0d0' | grep -v 'Average' > $ftmp_c0d0  
  ${ldavg_cmd} | grep ':' | grep -v 'ldavg' | grep -v 'Average' > $ftmp_ldavg  
  ${swp_cmd} | grep ':' | grep -v 'kswapd' | grep -v 'Average' > $ftmp_swp  
  ${eth_cmd} | grep 'lo' | grep -v 'Average' > $ftmp_lo  
  ${eth_cmd} | grep 'eth0' | grep -v 'Average' > $ftmp_eth0  
  ${sock_cmd} | grep ':' | grep -v 'totsck' | grep -v 'Average' > $ftmp_sock
```



```
cnt=`cat $ftmp_cpu | wc -l`  
i=1  
while [ ${i} -lt ${cnt} ]; do  
  ln1=`head -${i} $ftmp_cpu | tail -1`  
  ln_cpu=`echo $ln1 | awk '{print $4":"$5":"$6":"$7":"$8":"$4+$5+$6+$7+$8}'`  
  ln_mem=`head -${i} $ftmp_mem | tail -1 | awk '{print $5":"$9}'`  
  ln_io=`head -${i} $ftmp_io | tail -1 | awk '{print $5":"$6":"$9":"$11}'`  
  ln_io0=`head -${i} $ftmp_io0 | tail -1 | awk '{print $5":"$6":"$9":"$11}'`  
  ln_c0d0=`head -${i} $ftmp_c0d0 | tail -1 | awk '{print $5":"$6":"$9":"$11}'`  
  ln_ldavg=`head -${i} $ftmp_ldavg | tail -1 | awk '{print $5":"$6":"$7":"$3":"$4}'`  
  ln_swp=`head -${i} $ftmp_swp | tail -1 | awk '{print $4}'`  
  datetime=`echo $ln1 | awk '{print $1,$2}'`  
  cmd1="rrdtool update $rrd_db "  
  cmd2=`date -d "$date1 $datetime" +%s`  
  cmd3=`echo $ln_cpu":"$ln_mem":"$ln_io":"$ln_io0":"$ln_c0d0":"$ln_ldavg":"$ln_swp":"$ln_eth0":"$ln_lo":"$ln_sock`  
  cmd=`echo "$cmd1 $cmd2:$cmd3"`  
  $cmd  
  i=$((i + 1))  
done  
rm $ftmp_cpu $ftmp_mem $ftmp_io $ftmp_io0 $ftmp_c0d0 $ftmp_ldavg $ftmp_swp  
fi
```

Dane zgromadzone w pliku bazy rrd "rrd_db.rrd" mogą zostać wygenerowane w postaci ich reprezentacji graficznej za pomocą polecenia rrdtool graph. Parametry tego polecenia określają między innymi: źródło danych (czyli parametr który chcemy zobrazować), próbkę początkową i próbkę końcową, typy i kolory linii wykresu. Poniżej zamieszczono polecenia do generowania wykresów dla poszczególnych parametrów analizowanych w systemie monitoringu obciążenia serwerów Jagiellońskiej Biblioteki Cyfrowej, wraz z odpowiednimi przykładowymi wykresami.

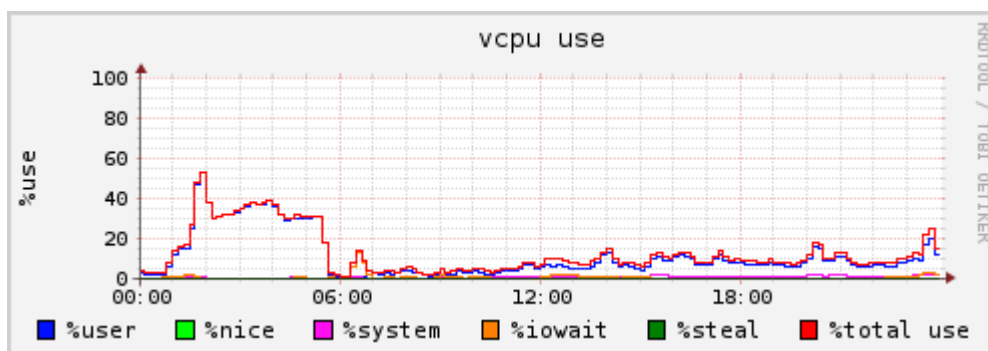
cpu.cfg

```
$rrd_cmd="graph - --imgformat=PNG \  
$start_time \  
$end_time \  
--title='vcpu use' \  
DEF:user=$RRD_DATA:c_user:AVERAGE \  
DEF:nice=$RRD_DATA:c_nice:AVERAGE \  
DEF:system=$RRD_DATA:c_system:AVERAGE \  
DEF:iowait=$RRD_DATA:c_iowait:AVERAGE \  
DEF:steal=$RRD_DATA:c_steal:AVERAGE \  
DEF:total=$RRD_DATA:c_total:AVERAGE \  

```

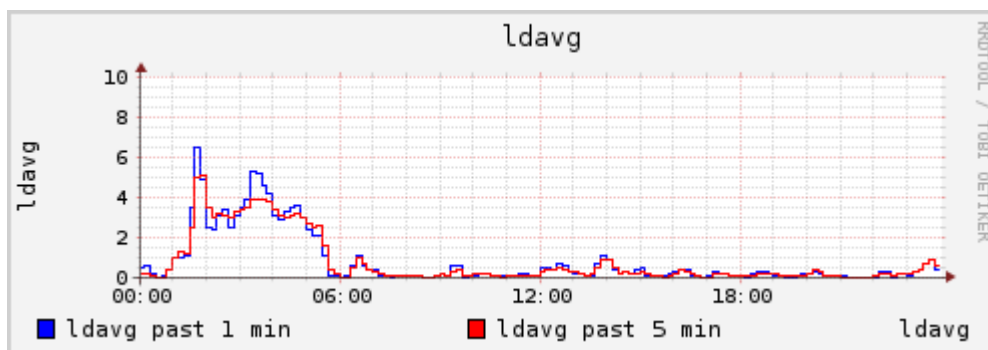



```
LINE1:user#000fff:%user' \  
LINE1:nice#00ff00:%nice' \  
LINE1:system#ff0ff0:%system' \  
LINE1:iowait#ff8000:%iowait' \  
LINE1:steal#008000:%steal' \  
LINE1:total#ff0000:%total use' \  
COMMENT:'cpu' \  
--vertical-label '%use' \  
";
```



Il. 1 Obciążenie wirtualnego procesora w okresie 24 godzinnym

```
ldavg.cfg  
$rrd_cmd="graph --imgformat=PNG \  
$start_time \  
$end_time \  
--title='ldavg' \  
DEF:ldavg1=$RRD_DATA:ldavg1:AVERAGE \  
DEF:ldavg5=$RRD_DATA:ldavg5:AVERAGE \  
DEF:ldavg15=$RRD_DATA:ldavg15:AVERAGE \  
LINE1:ldavg1#0000ff:'ldavg past 1 min' \  
LINE1:ldavg5#ff0000:'ldavg past 5 min' \  
COMMENT:'ldavg' \  
--vertical-label 'ldavg' \  
";
```

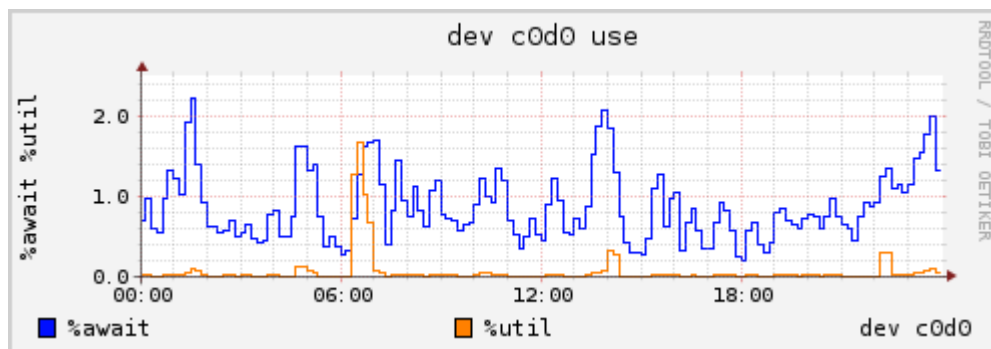


Il. 2 Średnie obciążenie systemu dla 1 i 5 minut w okresie 24 godzinnym



io_c0d0.cfg

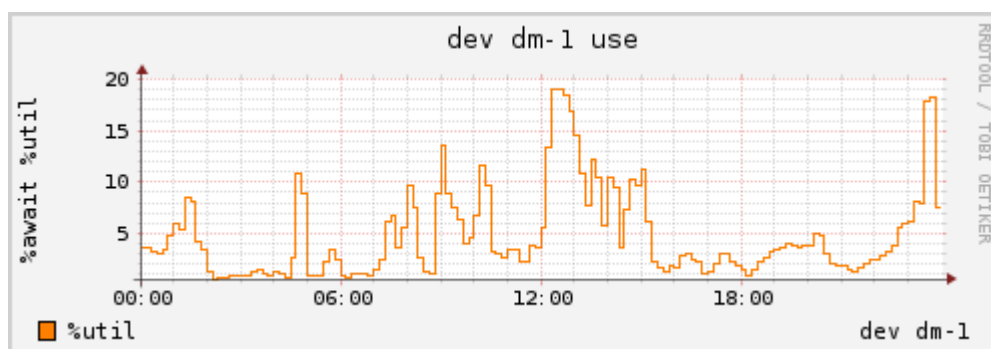
```
$rrd_cmd="graph --imgformat=PNG \  
$start_time \  
$end_time \  
--title='dev c0d0 use' \  
DEF:await=$RRD_DATA:c0d0_await:AVERAGE \  
DEF:util=$RRD_DATA:c0d0_util:AVERAGE \  
LINE1:await#000fff:%await' \  
LINE1:util#ff8000:%util' \  
COMMENT:'dev c0d0' \  
--vertical-label '%await %util' \  
";
```



Il. 3 Obciążenie dysków wewnętrznych w okresie 24 godzinnym

io_dm1.cfg

```
$rrd_cmd="graph --imgformat=PNG \  
$start_time \  
$end_time \  
--title='dev dm-1 use' \  
DEF:util=$RRD_DATA:i0_util:AVERAGE \  
LINE1:util#ff8000:%util' \  
COMMENT:'dev dm-1' \  
--vertical-label '%await %util' \  
";
```

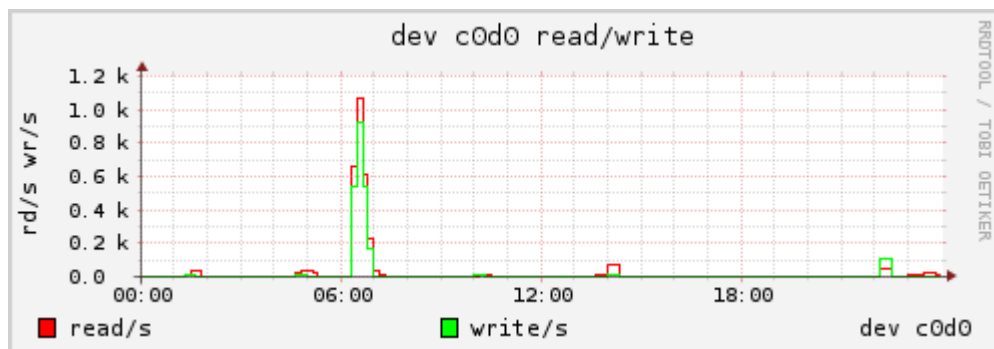


Il. 4 Obciążenie dysków wewnętrznych w okresie 24 godzinnym



rw_c0d0.cfg

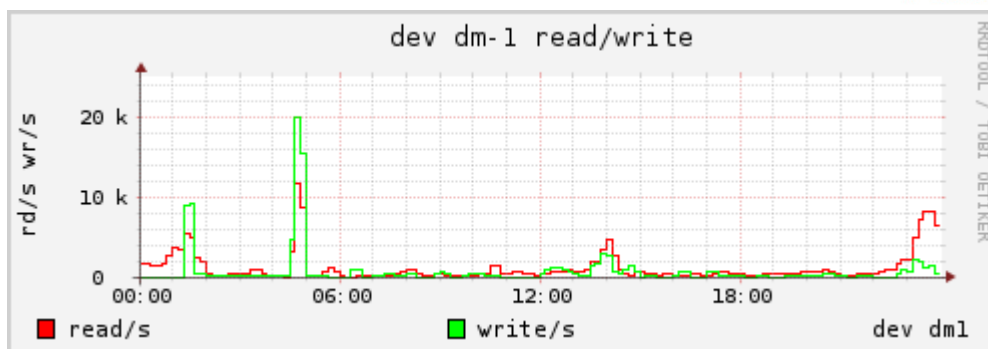
```
$rrd_cmd="graph - --imgformat=PNG \  
$start_time \  
$end_time \  
--title='dev c0d0 read/write' \  
DEF:rd_sec=$RRD_DATA:c0d0_rd_sec:AVERAGE \  
DEF:wr_sec=$RRD_DATA:c0d0_wr_sec:AVERAGE \  
LINE1:rd_sec#ff0000:'read/s' \  
LINE1:wr_sec#00ff00:'write/s' \  
COMMENT:'dev c0d0' \  
--vertical-label 'rd/s wr/s' \  
";
```



Il. 5 Obciążenie dysków wewnętrznych w funkcji zapis/odczyt w okresie 24 godzinnym

rw_dm1.cfg

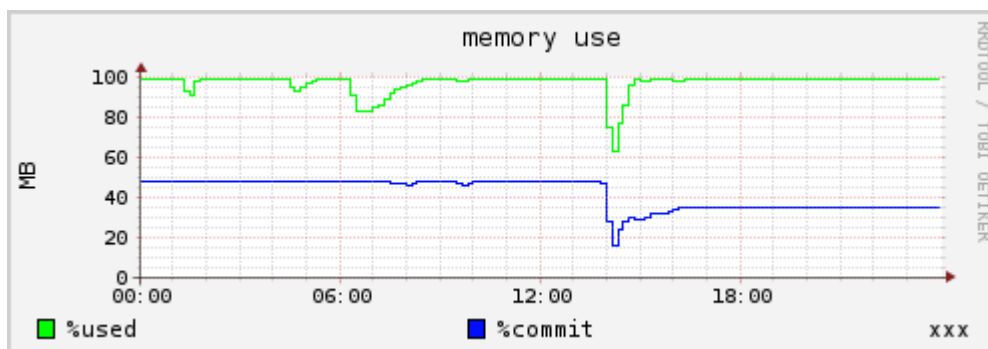
```
$rrd_cmd="graph - --imgformat=PNG \  
$start_time \  
$end_time \  
--title='dev dm-1 read/write' \  
DEF:rd_sec=$RRD_DATA:i0_rd_sec:AVERAGE \  
DEF:wr_sec=$RRD_DATA:i0_wr_sec:AVERAGE \  
LINE1:rd_sec#ff0000:'read/s' \  
LINE1:wr_sec#00ff00:'write/s' \  
COMMENT:'dev dm-1' \  
--vertical-label 'rd/s wr/s' \  
";
```



Il. 6 Obciążenie dysków wewnętrznych w funkcji zapis/odczyt w okresie 24 godzinnym

mem.cfg

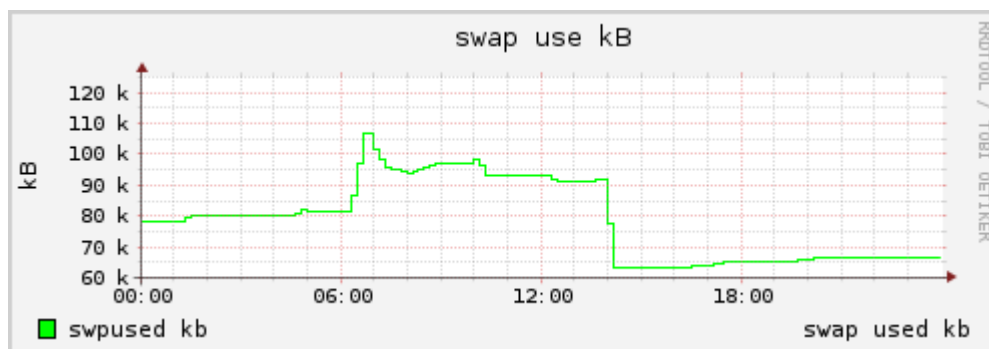
```
$rrd_cmd="graph --imgformat=PNG \  
$start_time \  
$end_time \  
--title='memory use' \  
DEF:memused=$RRD_DATA:m_memused:AVERAGE \  
DEF:commit=$RRD_DATA:m_commit:AVERAGE \  
LINE1:memused#00ff00:%used' \  
LINE1:commit#0000ff:%commit' \  
--vertical-label='MB' \  
";
```



Il. 7 Obciążenie pamięci w okresie 24 godzinnym

swap.cfg

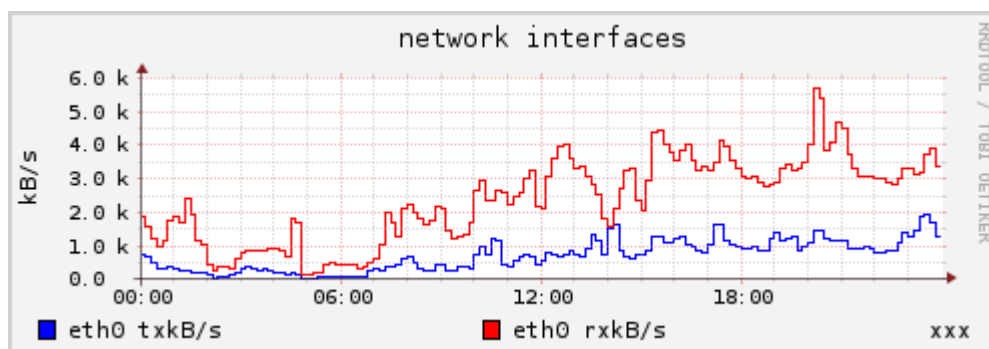
```
$rrd_cmd="graph --imgformat=PNG \  
$start_time \  
$end_time \  
--title='swap use kB' \  
DEF:kbswpused=$RRD_DATA:kbswpused:AVERAGE \  
LINE1:kbswpused#00ff00:'swpused kb' \  
COMMENT:'swap used kb' \  
--vertical-label='kB' \  
";
```



Il. 8 Obciążenie dodatkowej wolnej przestrzeni dyskowej w okresie 24 godzinny

net.cfg

```
$rrd_cmd="graph --imgformat=PNG \  
$start_time \  
$end_time \  
--title='network interfaces' \  
DEF:eth0_rxB=$RRD_DATA_NET:eth0_rxB:AVERAGE \  
DEF:eth0_txB=$RRD_DATA_NET:eth0_txB:AVERAGE \  
DEF:lo_rxB=$RRD_DATA_NET:lo_rxB:AVERAGE \  
DEF:lo_txB=$RRD_DATA_NET:lo_txB:AVERAGE \  
LINE1:eth0_txB#0000ff:eth0 txB/s' \  
LINE1:eth0_rxB#ff0000:eth0 rxB/s' \  
--vertical-label='kB/s' \  
";
```



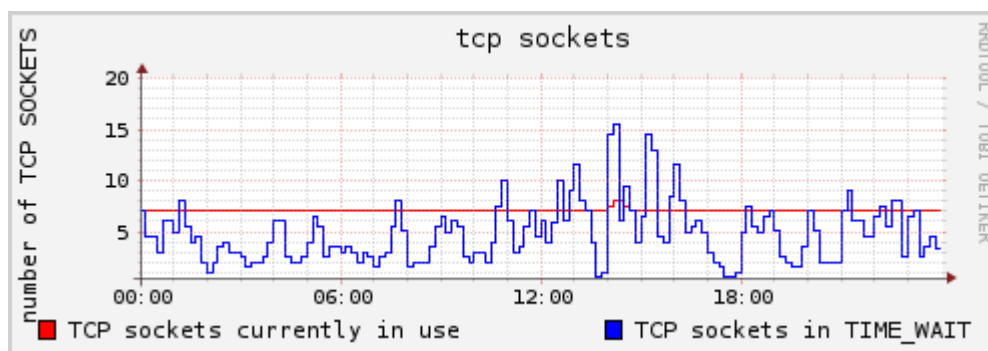
Il. 9 Obciążenie sieci w okresie 24 godzinny

sock.cfg

```
$rrd_cmd="graph --imgformat=PNG \  
$start_time \  
$end_time \  
--title='tcp sockets' \  
DEF:tcpsock=$RRD_DATA_NET:tcpsock:AVERAGE \  
";
```



```
DEF:tcp_tw=$RRRD_DATA_NET:tcp_tw:AVERAGE \  
LINE1:tcpsock#ff0000:'TCP sockets currently in use' \  
LINE1:tcp_tw#0000ff:'TCP sockets in TIME_WAIT' \  
COMMENT:'number of TCP SOCKETS' \  
--vertical-label='number of TCP SOCKETS' \  
";
```



Il. 10 Obciążenie gniazd sieciowych w okresie 24 godzinnym

2.2. Dostosowywanie konfiguracji systemu dLibra Jagiellońskiej Biblioteki Cyfrowej - rozdział serwisów dLibry na oddzielne serwery

System dLibra jest złożony z aplikacji serwera (dLibra – Serwer) oraz aplikacji klienckich (dLibra – Redaktor , dLibra - Czytelnik).

dLibra – Redaktor - aplikacja kliencka instalowana i działająca na komputerze użytkownika, jest przeznaczona do realizacji zadań związanych z tworzeniem, zarządzaniem i utrzymaniem biblioteki cyfrowej (dodawanie publikacji, wprowadzanie i modyfikacja metadanych itp.).

dLibra – Czytelnik – aplikacja przeznaczona do udostępniania zawartości biblioteki cyfrowej użytkownikowi końcowemu – czytelnikowi, jest oparta o serwer www Apache Tomcat.

dLibra – Serwer – serwer systemu dLibra jest złożony z szeregu współdziałających ze sobą usług sieciowych, z których każda realizuje odpowiednią funkcjonalność systemu. W skład serwera dLibra wchodzi następujące usługi:

- **Content Server** - odpowiada za przechowywanie i udostępnianie treści obiektów umieszczanych w bibliotece cyfrowej.
- **Metadata Server** - odpowiada za przechowywanie i udostępnianie metadanych na temat obiektów umieszczonych w bibliotece cyfrowej. Zarządza również wewnętrzną strukturą biblioteki, podziałem na katalogi, kolekcje, zestawem atrybutów, słownikami wartości atrybutów itd.



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

- **Search Server** - odpowiedzialny za indeksowanie i przeszukiwanie treści oraz metadanych na temat obiektów dostępnych w bibliotece cyfrowej.
- **Distributed Search Server** - odpowiedzialny za pobieranie, przechowywanie i udostępnianie metadanych na temat obiektów dostępnych w innych bibliotekach cyfrowych dostępnych poprzez protokół OAI-PMH.
- **User Server** - przechowuje informacje na temat użytkowników, grup użytkowników oraz uprawnień do poszczególnych publikacji. Jest wykorzystywany do autentykacji i autoryzacji przy dostępie do zasobów biblioteki cyfrowej.
- **Event Server** - wykorzystywany do asynchronicznej komunikacji pomiędzy pozostałymi serwerami systemu dLibra.
- **System Services** - usługa wykorzystywana do łączenia usług systemu dLibra w bibliotekę cyfrową oraz do autoryzacji dostępu pomiędzy usługami.

Podstawowy wariant konfiguracji systemu dLibra to wszystkie usługi uruchamiane na pojedynczym serwerze jako jedna grupa. W przypadku potrzeby rozbudowy biblioteki cyfrowej istnieje możliwość przenoszenia poszczególnych komponentów systemu na odrębne serwery. Każda z usług systemu dLibra może być uruchomiona na osobnym serwerze lub też usługi te mogą być łączone w grupy.

Aspektami, które należy uwzględnić przy analizowaniu potencjalnego obciążenia generowanego przez oprogramowanie dLibra są przede wszystkim:

- średnia liczba równoczesnych czytelników,
- łączna liczba dostępnych publikacji,
- średnia liczba nowych publikacji w ciągu tygodnia/dnia (ze szczególnym uwzględnieniem publikacji pełnotekstowych),
- średni rozmiar publikacji (liczba plików, liczba bajtów).

Początkowo system dLibra Jagiellońskiej Biblioteki Cyfrowej został uruchomiony w podstawowym wariantcie konfiguracji systemu dLibra, czyli wszystkie usługi dLibry działały na pojedynczym serwerze jako jedna grupa.

Jednakże taka konfiguracja nie dawała możliwości określenia wpływu poszczególnych serwisów dLibry na całkowite obciążenie serwera, oraz dawała potencjalną możliwość zawłaszczenia dostępnych zasobów sprzętowych przez pojedynczy serwis, co w konsekwencji mogło prowadzić do niedostępności pozostałych serwisów dLibry. Dlatego zdecydowano o rozdzieleniu wybranych serwisów dLibry i uruchomieniu ich na oddzielnych serwerach. Ze strony infrastruktury serwerowej do dyspozycji były dwa serwery fizyczne w następujących konfiguracjach:

- **serwer HP DL 360 G6**
1 x 4core CPU Intel Xeon E5520 2.27 GHz
RAM 8GB
- **serwer HP DL 380 G6**
2 x 4core CPU Intel Xeon E5520 2.27 GHz
RAM 24GB

Bazując na powyższej infrastrukturze serwerowej oraz na informacjach na temat funkcji i specyfiki poszczególnych serwisów systemu dLibra opisanych w dokumentacji (Strona domowa systemów dArceo, dLab, dLibra, dMuseion, 2013) w przypadku Jagiellońskiej Biblioteki Cyfrowej zdecydowano o rozdzieleniu serwisów dLibry w następujący sposób:

- Content Server, Metadata Server, Search Server (z wyłączeniem usługi indeksowania Index Server), User Server, Event Server, System Services
- Index Server (usługa indeksowania wydzielona z Search Server)
- dLibra – czytelnik (serwer www Apache Tomcat)
- Database Server (serwer bazy danych dla serwisów dLibry)

Rozdział powyższych serwisów w odniesieniu do fizycznych zasobów sprzętowych został zrealizowany w sposób następujący:

- Serwisy: Content Server, Metadata Server, Search Server (z wyłączeniem usługi indeksowania Index Server), User Server, Event Server, System Services zostały uruchomione na fizycznym serwerze **HP DL 360 G6** z systemem operacyjnym Linux Debian.
- Na serwerze **HP DL 380 G6** został zainstalowany system Vmware ESXi, na którego bazie zostały utworzone oddzielne maszyny wirtualne dedykowane odpowiednio dedykowane dla pozostałych usług systemu dLibra (Index Server, dLibra – czytelnik, Database Server).

Każdy z wydzielonych serwerów usług systemu dLibra Jagiellońskiej Biblioteki Cyfrowej został objęty opisanym systemem monitoringu wykorzystania zasobów. W celu ułatwienia korzystania z systemów monitoringu poszczególnych serwerów został stworzony jeden wspólny interfejs tego systemu, obejmujący wszystkie monitorowane serwery.

Przydział fizycznych zasobów serwera dla maszyn wirtualnych został dobrany na podstawie długookresowej analizy statystyk z systemu monitoringu wykorzystania zasobów. W chwili obecnej przydział ten wygląda następująco:



UNIwersytet
JAGIELLOŃSKI
W KRAKOWIE

- **Index Server:**

CPU	1 core of CPU Intel Xeon E5520 2.27 GHz
RAM	2,2 GB
Disk Space	50 GB

- **Database Server:**

CPU	4 core of CPU Intel Xeon E5520 2.27 GHz
RAM	8 GB
Disk Space	20 GB

- **dLibra – czytelnik (www Apache Tomcat):**

CPU	1 core of CPU Intel Xeon E5520 2.27 GHz
RAM	4 GB
Disk Space	20 GB

3. Protokół OAI-PMH

Podstawowym protokołem wymiany danych, używanym do importu i eksportu metadanych w Jagiellońskiej Bibliotece Cyfrowej jest OAI-PMH.

OAI (Open Archives Initiative) to inicjatywa dążąca do umożliwienia współpracy między bibliotekami/repozytoriami cyfrowymi. PMH (Protocol for Metadata Harvesting) to protokół umożliwiający gromadzenie metadanych rekordów zawartych w różnych repozytoriach.

Protokół PMH powstał w 1999 roku jako sposób na wymianę metadanych między repozytoriami, co ma z kolei umożliwić równoczesne wyszukiwanie w wielu repozytoriach. Metoda 'cross-search', polegająca na tym, że repozytoria nawzajem przekazują sobie żądania wyszukiwania i ich wyniki, została uznana za niewłaściwą z powodu wynikających z niej kłopotów ze skalowaniem wydajności. Zamiast tego zdecydowano się na tzw. „harvesting”, którego podstawową cechą jest podział na serwery danych i serwery gromadzące metadane. Według tej filozofii samo repozytorium nie musi umożliwiać wyszukiwania, musi za to odpowiadać na żądania pobierania metadanych przez serwer gromadzący metadane, tzw. harvester. Dopiero harvester udostępnia zgromadzone dane dla różnych usług, w tym usługi wyszukiwania.

W celu zapewnienia maksymalnej dostępności protokołu OAI-PMH, został on oparty na otwartym protokole przesyłania danych w sieci Internet – HTTP oraz otwartym protokole danych – XML.

Zapytania do repozytorium formułowane są jako zapytania GET lub POST protokołu HTTP, gdzie rodzaj i szczegóły zapytania stanowią część adresu URL.

Odpowiedzi na zapytania to strumień typu XML, otrzymany przez protokół HTTP. Oprócz nagłówka XML zawiera on sygnaturę czasową, ewentualny kod błędu oraz wynik zapytania. OAI-PMH wspiera dowolny XML-owy format metadanych, ale wymusza stosowanie formatu Dublin Core jako formatu bazowego, aby umożliwić współpracę wszystkich repozytoriów na podstawowym poziomie.



3.1. Zapytania protokołu OAI-PMH

Zapytania protokołu OAI-PMH formułowane są jako parametry HTTP dodawane do adresu URL interfejsu PMH, np.

<http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=Identify>

Wymagany jest przynajmniej jeden parametr 'verb', dodatkowe parametry zależą od rodzaju zapytania.

1. Identify

Zwraca podstawowe informacje o repozytorium.

np.

<http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=Identify>

2. ListMetadataFormats

Zwraca formaty metadanych obsługiwane przez repozytorium.

np.

<http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=ListMetadataFormats&identifier=oai:jbc.bj.uj.edu.pl:23520>

3. ListSets

Zwraca strukturę zbiorów repozytorium – w przypadku bibliotek opartych na systemie dLibra jest to struktura zdefiniowanych kolekcji.

np.

[http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?
verb=ListSets](http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=ListSets)

4. ListIdentifiers

Pobiera nagłówki rekordów z repozytorium.

np.

[http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?
verb=ListIdentifiers&
metadataPrefix=oai_dc](http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=ListIdentifiers&metadataPrefix=oai_dc)

5. ListRecords

Pobiera rekordy z repozytorium.

np.

[http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?
verb=ListRecords&
metadataPrefix=oai_dc](http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=ListRecords&metadataPrefix=oai_dc)

6. GetRecord

Pobiera pojedynczy rekord z repozytorium.

np.

```
http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?  
verb=GetRecord&  
metadataPrefix=oai_dc&  
identifier=oai:jbc.bj.uj.edu.pl:23520
```

3.2. Odpowiedzi protokołu OAI-PMH

Odpowiedzi protokołu OAI-PMH przesyłane są protokołem HTTP jako dane typu text/xml.

Odpowiedź to poprawnie sformułowany XML rozpoczynający się od deklaracji:

```
<?xml version="1.0" encoding="UTF-8" ?>
```

zawierający element główny OAI-PMH z atrybutami:

```
xmlns, xmlns:xsi, xsi:schemaLocation
```

i trzema elementami potomnymi:

```
responseDate (zawierającym sygnaturę czasową), request (zawierającym oryginalne zapytanie)
```

oraz elementem właściwym dla zapytania lub zgłoszeniem błędu.

Przykładowo, odpowiedź na zapytanie:

```
http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=Identify
```

wygląda następująco:

```
<?xml version="1.0" encoding="utf-8"?>  
<?xml-stylesheet type="text/xsl" href="http://jbc.bj.uj.edu.pl/style/common/xsl/oai-style.xsl"?>  
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/  
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">  
  <responseDate>2013-06-18T12:44:12Z</responseDate>  
  <request verb="Identify">  
    http://jbc.bj.uj.edu.pl/oai-pmh-repository.xml</request>  
  <Identify>  
    <repositoryName><![CDATA[Jagiellonian Digital Library]]</repositoryName>  
  <baseURL>http://jbc.bj.uj.edu.pl/oai-pmh-repository.xml</baseURL>  
  <protocolVersion>2.0</protocolVersion>  
  <adminEmail>dlibra-techhelp@man.poznan.pl</adminEmail>
```



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

```
<earliestDatestamp>2008-11-19T11:20:44Z</earliestDatestamp>
<deletedRecord>persistent</deletedRecord>
<granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
<description>
  <friends
    xmlns="http://www.openarchives.org/OAI/2.0/friends/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/friends/
      http://www.openarchives.org/OAI/2.0/friends.xsd">
<baseURL>http://dlibra.psnk.pl/biblioteka/dlibra/oai-pmh-repository.xml</baseURL>
    </friends>
  </description>
  <description>
    <oai-identifier xmlns="http://www.openarchives.org/OAI/2.0/oai-identifier"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier
        http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
      <scheme>oai</scheme>
      <repositoryIdentifier>jbc.bj.uj.edu.pl</repositoryIdentifier>
      <delimiter>:</delimiter>
      <sampleIdentifier>oai:jbc.bj.uj.edu.pl:1234</sampleIdentifier>
    </oai-identifier>
  </description>
  <description>
    <toolkit xmlns="http://oai.dlib.vt.edu/OAI/metadata/toolkit"
      xsi:schemaLocation="http://oai.dlib.vt.edu/OAI/metadata/toolkit
        http://jbc.bj.uj.edu.pl/toolkit.xsd">
      <title>dLibra Digital Library Framework</title>
      <author>
        <email>dlibra@man.poznan.pl</email>
        <institution>PSNC (http://www.man.poznan.pl)</institution>
      </author>
      <version>5.2.0</version>
      <toolkitIcon>http://dlibra.psnk.pl/images/dlibra_stamp.jpg</toolkitIcon>
      <URL>http://dlibra.psnk.pl/</URL>
    </toolkit>
  </description>
</Identify>
</OAI-PMH>
```

3.3. Format metadanych Dublin Core

Dla zapytań, które zwracają metadane obiektu (ListRecords i GetRecord) wynikowe dane XML będą zawierały osadzone metadane w określonym formacie, którym najczęściej będzie Dublin Core.

Dublin Core to format metadanych dążący do stworzenia podstawowego opisu dla zasobów. Może być używany do opisu obiektów fizycznych, takich jak książki czy dzieła sztuki, ale najczęściej używany jest dla zasobów dostępnych w Internecie: tekstów, obrazów, filmów, muzyki. W swojej podstawowej wersji składa się z 15 następujących elementów.

1. Title/Tytuł
2. Creator/Autor
3. Subject/Temat i słowa kluczowe
4. Description/Opis
5. Publisher/Wydawca
6. Contributor/Współtwórca
7. Date/Data wydania
8. Type/Typ zasobu
9. Format/Format
10. Identifier/Identyfikator zasobu
11. Source/Źródło
12. Language/Język
13. Relation/Powiązania
14. Coverage/Zakres
15. Rights/Prawa

W Jagiellońskiej Bibliotece Cyfrowej format Dublin Core został bardzo mocno rozbudowany i w obecnym momencie obejmuje 19 atrybutów głównych i 25 podatrybutów. Związane to było z wprowadzaniem do JBC nowych typów dokumentów, które do dokładniejszego opisu wymagały zastosowania nowych atrybutów.



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

3.4. Pobieranie metadanych z systemu dLibra za pomocą protokołu OAI-PMH

Jagiellońska Biblioteka Cyfrowa, prezentująca Bazę Wiedzy UJ (kolekcja: „Projekty – Synat”) w Internecie, oparta jest na systemie dla bibliotek cyfrowych o nazwie dLibra. Platforma ta od wersji 2.0 obsługuje udostępnianie metadanych repozytorium przez protokół OAI-PMH. Adres interfejsu OAI-PMH dla biblioteki opartej na dLibrze to:

<http://<biblioteka cyfrowa>/dlibra/oai-pmh-repository.xml>

czyli w przypadku Jagiellońskiej Biblioteki Cyfrowej:

<http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml>

Pobieranie metadanych z JBC trzeba zacząć od pobrania listy kolekcji:

<http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?>

verb=ListSets

co daje w wyniku następujący plik XML:

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-stylesheet type="text/xsl" href="http://jbc.bj.uj.edu.pl/style/common/xsl/oai-style.xsl"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2013-06-18T13:50:03Z</responseDate>
  <request
resumptionToken="B60329F2398B9EF59314C30354BF903EListSets1371556112844_DL_LAST_ITEM_50" verb="ListSets">
    http://jbc.bj.uj.edu.pl/oai-pmh-repository.xml</request>
  <ListSets>
    <set>
    <setSpec>dLibraDigitalLibrary:UJ:WPiA:KHiDPiP</setSpec>
      <setName><![CDATA[Katedra Historii Doktryn Politycznych i Prawnych]]</setName>
      <setDescription>
        <oai_dc:dc
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:description xml:lang="pl"><![CDATA[<a
href="http://www.law.uj.edu.pl/~khdpip">www.law.uj.edu.pl/~khdpip</a>]]</dc:description>
          <dc:description xml:lang="en"><![CDATA[<a
href="http://www.law.uj.edu.pl/~khdpip">www.law.uj.edu.pl/~khdpip</a>]]</dc:description>
        </oai_dc:dc>
```



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

```
</setDescription>
</set>
<set>
<setSpec>dLibraDigitalLibrary:UJ:WPiA:KHiDPiP:criteria</setSpec>
<setName><![CDATA[Katedra Historii Doktryn Politycznych i Prawnych (Gateway for dynamic
set)]]></setName>
<setDescription>
<oai_dc:dc
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:description xml:lang="pl"><![CDATA[<a href="http://www.law.uj.edu.pl/~khdpip">www.law.uj.edu.pl/~khdpip</a>
(Gateway for dynamic set)]]></dc:description>
<dc:description xml:lang="en"><![CDATA[<a
href="http://www.law.uj.edu.pl/~khdpip">www.law.uj.edu.pl/~khdpip</a> (Gateway for dynamic set)]]></dc:description>
</oai_dc:dc>
</setDescription>
</set>
(...)
<set>
<setSpec>criteria</setSpec>
<setName><![CDATA[(Gateway for dynamic set)]]></setName>
</set>
<resumptionToken completeListSize="452" cursor="50" expirationDate="2013-06-
18T14:20:03Z">B60329F2398B9EF59314C30354BF903EListSets1371556112844_DL_LAST_ITEM_100</resumptionToken>
</ListSets>
</OAI-PMH>
```

W pliku tym można wyróżnić nagłówki, osadzone w elemencie ListSets informacje o kolekcjach oraz tzw. resumptionToken. Ponieważ protokół OAI-PMH wspiera „stronicowanie” danych, w jednej odpowiedzi jest podawana ograniczona liczba rekordów – aby uzyskać ciąg dalszy, należy wystosować nowe zapytanie, podając otrzymany we wcześniejszym zapytaniu resumptionToken:

```
http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=ListSets&
resumptionToken=B60329F2398B9EF59314C30354BF903EListSets1371556112844_DL_LAST_ITEM_50
```




UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

Kontynuując w ten sposób, możemy zebrać komplet informacji na temat kolekcji określonych w Jagiellońskiej Bibliotece Cyfrowej. Nas interesują szczegóły dotyczące kolekcji „Projekty – Synat”:

```
<set>
  <setSpec>dLibraDigitalLibrary:Projekty:Synat</setSpec>
  <setName><![CDATA[Synat]]></setName>
</set>
```

Znaleziony w ten sposób identyfikator kolekcji dLibraDigitalLibrary:Projekty:Synat wykorzystamy do pobrania metadanych rekordów należących do tej kolekcji – za pomocą zapytania ListRecords do interfejsu OAI-PMH. Zapytanie takie wymaga obowiązkowo określenia także formatu metadanych – w naszym przypadku: oai_dc:

```
http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=ListRecords&
metadataPrefix=oai_dc&
set=dLibraDigitalLibrary:Projekty:Synat
```

Wynik zapytania jest następujący:

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-stylesheet type="text/xsl" href="http://jbc.bj.uj.edu.pl/style/common/xsl/oai-style.xsl"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2013-06-18T14:12:53Z</responseDate>
  <request metadataPrefix="oai_dc" verb="ListRecords" set="dLibraDigitalLibrary:Projekty:Synat">
http://jbc.bj.uj.edu.pl/oai-pmh-repository.xml</request>
  <ListRecords>
    <record>
      <header>
        <identifier>oai:jbc.bj.uj.edu.pl:203041</identifier>
        <timestamp>2012-12-21T11:34:18Z</timestamp>
        <setSpec>dLibraDigitalLibrary:KlasyfikacjaSystematyczna:Biologia</setSpec>
<setSpec>dLibraDigitalLibrary:KlasyfikacjaSystematyczna:Biologia:Weterynaria</setSpec>
<setSpec>dLibraDigitalLibrary:Czytelnie</setSpec>
<setSpec>dLibraDigitalLibrary:KlasyfikacjaSystematyczna</setSpec>
```



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

```
<setSpec>dLibraDigitalLibrary:KolekcjeTematyczne</setSpec>
<setSpec>dLibraDigitalLibrary:KolekcjeTematyczne:synat_inib</setSpec>
<setSpec>dLibraDigitalLibrary:Projekty:Synat</setSpec> <setSpec>dLibraDigitalLibrary</setSpec>
<setSpec>dLibraDigitalLibrary:Projekty</setSpec> <setSpec>dLibraDigitalLibrary:Czytelnie:online</setSpec>
</header>

    <metadata>
      <oai_dc:dc
        xmlns:dc="http://purl.org/dc/elements/1.1/"
        xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
        <dc:title xml:lang="pl"><![CDATA[Baza Publikacji Pracowników Uniwersytetu Przyrodniczego w Lublinie]]></dc:title>
        <dc:subject xml:lang="pl"><![CDATA[medycyna weterynaryjna]]></dc:subject>
        <dc:subject xml:lang="pl"><![CDATA[bibliografia publikacji pracowników uczelni wyższej]]></dc:subject>
        <dc:description xml:lang="pl"><![CDATA[Biblioteka Główna Uniwersytetu Przyrodniczego w
Lublinie]]></dc:description>
        <dc:description xml:lang="pl"><![CDATA[SYZYF]]></dc:description>
        <dc:publisher xml:lang="pl"><![CDATA[Biblioteka Główna Uniwersytetu Przyrodniczego w Lublinie]]></dc:publisher>
        <dc:type xml:lang="pl"><![CDATA[strona naukowa]]></dc:type>
        <dc:format xml:lang="pl"><![CDATA[text/html]]></dc:format>
        <dc:identifier><![CDATA[http://jbc.bj.uj.edu.pl/Content/203041]]></dc:identifier>
        <dc:source
          xml:lang="pl"><![CDATA[http://syzyf1.ar.lublin.pl/~prace/public/index.php?config=jedn;
wyd_id=6]]></dc:source>
        <dc:language xml:lang="pl"><![CDATA[pol]]></dc:language>
        <dc:relation><![CDATA[oai:jbc.bj.uj.edu.pl:publication:214401]]></dc:relation>
      </oai_dc:dc>
    </metadata>
  </record>
</record>
<header>
  <identifier>oai:jbc.bj.uj.edu.pl:237822</identifier>
  <timestamp>2013-05-07T16:04:29Z</timestamp>
  <setSpec>dLibraDigitalLibrary:KlasyfikacjaSystematyczna:Biologia</setSpec>
<setSpec>dLibraDigitalLibrary:KlasyfikacjaSystematyczna</setSpec>
<setSpec>dLibraDigitalLibrary:Projekty:Synat</setSpec> <setSpec>dLibraDigitalLibrary</setSpec>
<setSpec>dLibraDigitalLibrary:Projekty</setSpec>
<setSpec>dLibraDigitalLibrary:KlasyfikacjaSystematyczna:Biologia:Zoologia</setSpec> </header>
  <metadata>
    <oai_dc:dc
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
      <dc:title xml:lang="pl"><![CDATA[Sekcja Ornitologiczna Polskiego Towarzystwa Zoologicznego]]></dc:title>
      <dc:subject xml:lang="pl"><![CDATA[zoologia]]></dc:subject>
```



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

```
<dc:subject xml:lang="pl"><![CDATA[ornitologia]]></dc:subject>
<dc:subject xml:lang="pl"><![CDATA[ptaki]]></dc:subject>
<dc:subject xml:lang="pl"><![CDATA[Polskie Towarzystwo Zoologiczne]]></dc:subject>
<dc:subject xml:lang="pl"><![CDATA[Polska]]></dc:subject>
<dc:description xml:lang="pl"><![CDATA[Zoologia]]></dc:description>
<dc:publisher xml:lang="pl"><![CDATA[Sekcja Ornitologiczna Polskiego Towarzystwa
Zoologicznego]]></dc:publisher>
<dc:date xml:lang="pl"><![CDATA[cop. 2009-2013]]></dc:date>
<dc:format xml:lang="pl"><![CDATA[text/html]]></dc:format>
<dc:format xml:lang="en"><![CDATA[text/html]]></dc:format>
<dc:identifier><![CDATA[http://jbc.bj.uj.edu.pl/Content/237822]]></dc:identifier>
<dc:source xml:lang="pl"><![CDATA[http://soptzool.wnb.uz.zgora.pl/]]></dc:source>
<dc:language xml:lang="pl"><![CDATA[pol]]></dc:language>
<dc:relation><![CDATA[oai:jbc.bj.uj.edu.pl:publication:249621]]></dc:relation>
</oai_dc:dc>

</metadata>
</record>

(...)
<resumptionToken completeListSize="4486" cursor="0" expirationDate="2013-06-
18T14:42:54Z">BD7D52421B06FA27D01F4CD54272C583ListRecords1371557574456_DL_LAST_ITEM_20_DL_METADATA
_oai_dc</resumptionToken> </ListRecords>
</OAI-PMH>
```

Jak widać, po nagłówku XML w pliku następuje element ListRecords, w którym osadzone są metadane rekordów w formacie DublinCore. Dodatkowo plik kończy się podaniem resumptionToken, którego iteracyjne stosowanie pozwala pobrać komplet danych:

```
http://jbc.bj.uj.edu.pl/dlibra/oai-pmh-repository.xml?verb=ListRecords&
resumptionToken=BD7D52421B06FA27D01F4CD54272C583ListRecords1371557574456_DL_LAST_ITEM_20_DL_M
ETADATA_oai_dc
```

W ten sposób otrzymane metadane zostały zaimportowane do portalu PASSIM, a wykorzystując dowolne funkcje importu z pliku XML w formacie DublinCore, można importować pełny lub okrojony zestaw metadanych do wszystkich baz wiedzy wykorzystujących protokół OAI-PMH.



4. Integracja aplikacji dLibra Jagiellońskiej Biblioteki Cyfrowej z Portalem UJ-BJ

W celu integracji dostępu do Jagiellońskiej Biblioteki Cyfrowej zarówno z portalu Biblioteki Jagiellońskiej jak i z głównego portalu BJ należy stworzyć punkt wzajemnych odniesień. Jest to metoda prymitywna, niemniej pełniąca swoją funkcję informacyjną.

Integracja aplikacji dLibra Jagiellońskiej Biblioteki Cyfrowej z Portalem UJ-BJ zakłada:

- wspólny styl interfejsu stron www (układ, kolorystyka)
- umieszczenie w Portalu UJ-BJ okna wyszukiwarki umożliwiającej wyszukiwanie w zasobach Jagiellońskiej Biblioteki Cyfrowej / bazy wiedzy UJ

Formularz wyszukiwania umożliwiający wyszukiwanie proste w Jagiellońskiej Bibliotece Cyfrowej powinien zawierać następujące pola i parametry:

```
<form
name="smallsearch"
id="smallsearch"
method="get"
action="http://jbc.bj.uj.edu.pl/dLibra/results?action=SearchAction&amp" >

<input type="hidden" name="action" value="SearchAction" />
<input type="hidden" name="skipSearch" value="true" />
<input type="hidden" name="mdirids" id="mdirids" value=""/>
<input type="hidden" name="server:type" id="server:type" value="both" />
<input type="hidden" name="tempQueryType" id="tempQueryType" value="-3" />
<input type="hidden" name="encode" id="encode" value="false" />
<input type="hidden" name="isExpandable" id="isExpandable" value="on" />
<input type="hidden" id="isRemote" name="isRemote" value="off"/>
<input name="roleId" id="roleId" type="hidden"/>
<input name="queryType" id="queryType" type="hidden"/>
<input name="dirids" id="dirids" type="hidden" value="1"/>
<input name="rootid" id="rootid" type="hidden" value=""/>
<input type="hidden" name="localQueryType" id="localQueryType" value="-3">
<input type="hidden" name="remoteQueryType" id="remoteQueryType" value="-2" />
<input
type="text"
name="query"
id="query"
value=""/>
```



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

```
<input  
type="submit"  
value="Szukaj"  
id="search" />  
</form>
```

Mówiąc o prostocie tej metody nie podważa się tu absolutnie jej zalet. Natomiast obserwując rozwiązania światowe czy choćby krajowe jak przedstawione na il. 11 i 12 rozwiązanie z Biblioteki Uniwersytetu Adama Mickiewicza w Poznaniu, widać tendencję do integracji usług i opracowaniem interfejsu jak najbardziej zbliżonemu do Google.

[Elektroniczny katalog kartkowy](#) | [Libsmart copy](#) | [Portale dziedzinowe](#) | [Pomoc](#) | [O Summonie](#) | [Opinie](#) | polski ▾

Odkryj świat zasobów biblioteki

Wyszukiwanie zaawansowane

Il. 11. Widok strony multiwyszukiwarki Summon i rodzajów zasobów jakie integruje na UAM w Poznaniu

Rozwiązania integrujące wszystkie możliwe serwisy w jednym oknie i dające listę wyników są najbardziej pożądanymi przez użytkowników o czym świadczy sukces wyszukiwarki Google wśród szerokiej rzeszy Internautów, czy Google Scholar wśród naukowców.



MULTIWYSZUKIWARKA
KATALOG

Szukaj

INFORMATOR

- Godziny otwarcia, lokalizacja
- Czytelnie i pracownie
- Wypożyczalnia
- Wypożyczalnia międzybiblioteczna
- Informatorium
- Regulamin i opłaty
- Debiuty i spotkania

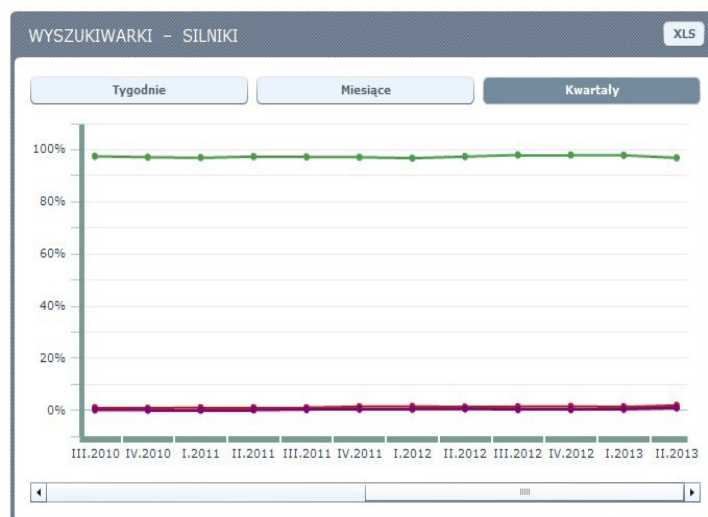
ZASOBY I ŹRÓDŁA

- Multiwyszukiwarka SUMMON
- Katalogi
- Bazy danych
- Książki elektroniczne
- Czasopisma elektroniczne
- Repozytorium AMUR
- Biblioteki wirtualne

SERWISY I USŁUGI

- Portale dziedzinowe
- Usługi online
- E-księgarnia
- Zaproponuj kupno książki
- Biblioteka poszukuje
- Oferta dubletów

II. 12 Strona główna Biblioteki UAM w Poznaniu



WYSZUKIWARKI - SILNIKI

Lp.	Nazwa	17.06-23.06.2013	18.06-24.06.2012	20.06-26.06.2011
1	Google	96.88%	97.60%	97.69%
2	MSN	1.85%	1.51%	0.98%
3	AskJeeves	1.00%	0.59%	0.15%
4	NetSprint	0.15%	0.17%	0.90%
5	Yahoo	0.12%	0.11%	0.24%

Znajdź:

II. 13 Ranking najpopularniejszych wyszukiwarek w Polsce

Bardzo dobrze charakteryzuje to wykres i tabela na il. 13 gdzie przedstawiono ranking silników wyszukiwawczych za ostatnie 2 lata. Ranking obejmuje polskie witryny do których internauci łączą się z obszaru Polski. Ranking jest tworzony na podstawie uśrednionego dla witryn udziału procentowego wizyt rozpoczętych wejściami z silników wyszukiwarek na serwisy objęte badaniem gemiusTraffic. Dominacja firmy Google, trwająca już od ponad dekady nie jest zagrożona w najmniejszym aspekcie, chyba że do głosu dojdą czynniki polityczne takie jak afera PRISM, które mogą podważyć zaufanie użytkowników do tego informacyjnego monopolisty. Niemniej jednak preferencje użytkowników są jasno określone poprzez sam fakt utrzymywania się wyszukiwarki Google jako narzędzia wykorzystwanego przez 97% badanych Internautów.

5. Wnioski

Sposób zarządzania Jagiellońską Biblioteką Cyfrową jest procesem dynamicznym i wciąż jest optymalizowany. Przyczynia się do tego permanentnie wzrastająca liczba publikacji z jednej strony, a użytkowników z drugiej. Pewne funkcjonalności w zależności od obciążenia są przenoszone pomiędzy serwerami, tak aby ich wydajność pozwalała na sprawną pracę redaktorów, a zarazem wyszukiwanie i ładowanie publikacji od strony użytkowników. Mankamentem jest brak dostępu do kodu źródłowego oprogramowania dLibra co bardzo spowalnia reakcję na zgłaszane przez redaktorów czy czytelników problemy. Rozwiązaniem może tu być wykupienie dostępu do kodu źródłowego i wspólne prace z PCSS nad udoskonaleniem oprogramowania, a w efekcie próbą jego zupełnego „uwolnienia”. Tylko wolnodostępna licencja pozwoli na dalszy efektywny jego rozwój. Pozostawienie go na poziomie oprogramowania komercyjnego skutecznie ograniczy jego rozwój w coraz szybciej zmieniających się warunkach technologicznej ekspansji.

Sposób przechowywania metadanych Jagiellońskiej Biblioteki Cyfrowej w bazie danych systemu dLibra pozwala na łatwe wyeksportowanie tych danych do innych systemów. Eksport taki można wykonać np. w formie xml, ale lepszym rozwiązaniem wydaje się użycie protokołu OAI-PMH, opartego na HTTP i XML. Jednakże przy próbie importu danych, dla obiektu przedstawionego na il. 14, po protokole OAI-PMH do platformy PASSIM kazało się że pewne metadane zostały obcięte.



Wydanie

- Opis
- Informacje
- Treść
- Treść (nowe okno)
- Pobierz
- Komentarz
- Podobne wydania

Eksport metadanych

- OAI-PMH
- RDF
- RIS
- BIBTEX

Ulubione pozycje

Dodaj do Ulubionych

Tagowanie

Zatwierdź

tylko prywatne

Zapisz ten adres...

Dodaj do "Zakładek"

Bookmark

Opis wydania

About : Chemistry

Tytuł:
About : Chemistry

Wydawca:
About Inc

Abstrakt:
Strona stanowi część sieci About.com. Strona daje dostęp do sieciowych zasobów z dziedziny chemii w Internecie. Zawiera linki do zasobów internetowych z około 20 obszarów tematycznych. Strona zawiera również aktualne wiadomości związane z badaniami chemicznymi, artykuły, czat i forum.

Data:
cop. 2011

Data dostępu:
2011

Język:
eng

Temat i słowa kluczowe:
chemia ; zasoby sieciowe ; katalogi zasobów sieciowych ; wyszukiwanie informacji

KBN:
Nauki chemiczne

UKD:
54 Chemia. Krystalografia. Mineralogia

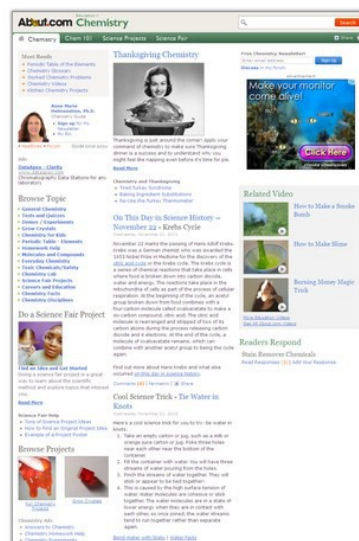
Format:
text/html

Źródło:
kliknij tutaj, żeby przejść

Identyfikator dokumentu cyfrowego:
DIGONLINE003269

Identyfikator zasobu:
oai:jbc.bj.uj.edu.pl:204242

Źródło finansowania:
Synat



Pokaż treść!

Prezentowany opis stanowi element sporządzonej na podstawie analizy krajowych i zagranicznych zasobów cyfrowych bazy wiedzy rejestrującej internetowe źródła informacji z zakresu nauk przyrodniczo-matematycznych. Został przygotowany przez pracowników Instytutu Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego w ramach projektu Synat. Treść publikacji stanowi odnośnik do opisanego zasobu (ten sam odnośnik znajduje się w polu Źródło). Całą bazę można przeglądać, wybierając w lewym panelu nawigacyjnym kolekcję tematyczną „Baza zasobów cyfrowych z zakresu nauk przyrodniczo-matematycznych”

Il. 14. Widok strony dla której był wykonywany eksport metadanych po protokole OAI-PMH do platformy PASSIM



UNIwersYTET
JAGIELLOŃSKI
W KRAKOWIE

Różnica pomiędzy prostym importem i przypisanym do niego zestawem metadanych, a pełnym zestawem dostępnym po utworzeniu źródła pliku została przedstawiona odpowiednio na il. 15 i 16.

• **oai:jbc.bj.uj.edu.pl:204242**

- o dc:title (pl) : About : Chemistry
- o dc:subject (pl) : chemia
- o dc:subject (pl) : zasoby sieciowe
- o dc:subject (pl) : katalogi zasobów sieciowych
- o dc:subject (pl) : wyszukiwanie informacji
- o dc:publisher (pl) : About Inc
- o dc:date (pl) : cop. 2011
- o dc:format (pl) : text/html
- o dc:identifier : http://jbc.bj.uj.edu.pl/Content/204242
- o dc:source (pl) : http://chemistry.about.com/
- o dc:language (pl) : eng
- o dc:relation : oai:jbc.bj.uj.edu.pl:publication:215618

Il. 15. Widok schematu metadanych wg protokołu OAI-PMH przy prostej formie importu

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <?xml-stylesheet type="text/xsl" href="http://jbc.bj.uj.edu.pl/style/common/xsl/oai-style.xsl"?>
3 <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
4   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
5   xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
6   http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
7   <responseDate>2013-07-02T11:24:47Z</responseDate>
8   <request identifier="oai:jbc.bj.uj.edu.pl:204242" metadataPrefix="dlibra_avs" verb="GetRecord">
9     http://jbc.bj.uj.edu.pl/oai-pmh-repository.xml</request>
10   <GetRecord>
11
12     <record>
13       <header>
14         <identifier>oai:jbc.bj.uj.edu.pl:204242</identifier>
15         <timestamp>2013-01-03T16:24:54Z</timestamp>
16         <setSpec>dlibraDigitalLibrary:Czytelnie</setSpec> <setSpec>dlibraDigitalLibrary:KlasyfikacjaSystemaryczna:Chemia</setSpec> <setSpec>dlibraDigitalLibrary:KlasyfikacjaSystemaryczna</setSpec>
17         <setSpec>dlibraDigitalLibrary:KolekcjeTematyczne</setSpec> <setSpec>dlibraDigitalLibrary:KolekcjeTematyczne:synat_inib</setSpec> <setSpec>dlibraDigitalLibrary</setSpec>
18         <setSpec>dlibraDigitalLibrary:Projekty</setSpec> <setSpec>dlibraDigitalLibrary:Projekty</setSpec> <setSpec>dlibraDigitalLibrary:Czytelnie</setSpec> </header>
19       <metadata>
20         <dlibra_avs:metadataDescription xmlns:dlibra_avs="http://jbc.bj.uj.edu.pl/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://jbc.bj.uj.edu.pl/ http://jbc.bj.uj.edu.pl/dlibra/attribute-schema.xsd">
21           <dlibra_avs:Title>
22             <value xml:lang="pl"><![CDATA[About : Chemistry]]></value>
23           </dlibra_avs:Title>
24           <dlibra_avs:Publisher>
25             <value xml:lang="pl"><![CDATA[About Inc]]></value>
26           </dlibra_avs:Publisher>
27           <dlibra_avs:Description>
28             <dlibra_avs:abstract xml:lang="pl"><![CDATA[Strona stanowi część sieci About.com. Strona daje dostęp do sieciowych zasobów z dziedziny chemii w Internecie. Zawiera linki do zasobów internetowych z około 20 obszarów tematycznych. Strona zawiera również aktualne wiadomości związane z badaniami chemicznymi, artykuły, czat i forum.]]></dlibra_avs:abstract>
29           </dlibra_avs:Description>
30           <dlibra_avs:Date>
31             <value xml:lang="pl"><![CDATA[cop. 2011]]></value>
32           </dlibra_avs:Date>
33           <dlibra_avs:access_date xml:lang="pl"><![CDATA[2011]]></dlibra_avs:access_date>
34           </dlibra_avs:Date>
35           <dlibra_avs:Language xml:lang="pl"><![CDATA[eng]]></dlibra_avs:Language>
36           <dlibra_avs:Subject xml:lang="pl"><![CDATA[chemia]]></dlibra_avs:Subject>
37           <dlibra_avs:Subject xml:lang="pl"><![CDATA[katalogi zasobów sieciowych]]></dlibra_avs:Subject>
38           <dlibra_avs:Subject xml:lang="pl"><![CDATA[wyszukiwanie informacji]]></dlibra_avs:Subject>
39           <dlibra_avs:Subject xml:lang="pl"><![CDATA[zasoby sieciowe]]></dlibra_avs:Subject>
40           <dlibra_avs:classifications>
41             <dlibra_avs:kbn xml:lang="pl"><![CDATA[Nauki chemiczne]]></dlibra_avs:kbn>
42             <dlibra_avs:ukd xml:lang="pl"><![CDATA[54 Chemia. Krystalografia. Mineralogia]]></dlibra_avs:ukd>
43           </dlibra_avs:classifications>
44           <dlibra_avs:degree>
45             <dlibra_avs:faculty/>
46           </dlibra_avs:degree>
47           <dlibra_avs:Format xml:lang="pl"><![CDATA[text/html]]></dlibra_avs:Format>
48           <dlibra_avs:Source>
49             <value xml:lang="pl"><![CDATA[http://chemistry.about.com/]]></value>
50           </dlibra_avs:Source>
51           <dlibra_avs:Relation/>
52           <dlibra_avs:Rights/>
53           <dlibra_avs:Identifier/>
54           <dlibra_avs:digitisationSponsor xml:lang="pl"><![CDATA[Synat]]></dlibra_avs:digitisationSponsor>
55           </dlibra_avs:DC>
56           <dlibra_avs:metadataDescription>
57         </dlibra_avs:metadataDescription>
58       </record> </GetRecord>
59 </OAI-PMH>
60
```

Il. 16. Widok pełnego zestawu metadanych po zastosowaniu podglądu źródła pliku importowanego przez protokół OAI-PMH z uwzględnieniem metadanych właściwych dla JBC

Różnica polegała na tym, że proste importowanie po protokole OAI-PMH nie uwzględniało zestawu atrybutów nadanych dodatkowo w JBC, a obejmowało tylko 15 podstawowych atrybutów formatu Dublin Core.

Takie przypadki pozwalają na lepsze poznanie oprogramowania i metod radzenia sobie z zaistniałymi, a nieprzewidywalnymi sytuacjami. Stąd koncepcja wykupienia dostępu do kodu źródłowego wydaje się być jedynym logicznym posunięciem w dalszym zarządzaniu oprogramowaniem dLibra.

Dostęp do kodu źródłowego ma również znaczenie przy integracji usług, bez niego sprzęgnięcie JBC z portalem UJ i BJ będzie wymagało dodatkowo nakładu pracy. Rozwijanie w przyszłości zintegrowanego wyszukiwania również byłoby łatwiejsze, gdyby BJ, a najlepiej wszyscy mieli dostęp do otwartego kodu źródłowego.

Bibliografia:

1. dLibra 5.0 – Instrukcja obsługi systemu. Rozdział 11. Komunikacja z innymi systemami informacyjnymi: 01. Protokół OAI-PMH (2010).
<https://confluence.man.poznan.pl/community/pages/viewpage.action?pageId=13631780>
[odczyt 26.06.2013]
2. dLibra – oficjalna strona projektu dLibra (2008). <http://dLibra.psnc.pl> [odczyt 26.06.2013]
3. dLibra – oficjalna strona społeczności projektu dLibra (2008).
<http://dLibra.psnc.pl/community> [odczyt 25.06.2013]
4. dLibra – forum użytkowników (2008). <http://dLibra.psnc.pl/forum> [odczyt 26.06.2013]
5. Dublin Core (2009). <http://dublincore.org> [odczyt 27.06.2013]
6. Methodology. Ranking Web of Repositories (2012).
<http://repositories.webometrics.info/en/Methodology> [odczyt 27.06.2013]
7. OAI for Beginners - the Open Archives Forum online tutorial (2012)
<http://www.oaforum.org/tutorial/english/intro.htm> [odczyt 25.06.2013]
8. Open Archives Initiative (2012). <http://www.openarchives.org> [odczyt 27.06.2013]