

39017806021

Jelenić, Ivan

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Pula / Sveučilište Jurja Dobrile u Puli**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:999:246306>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-01-18**



Repository / Repozitorij:

[The Repository of the everything and nothing](#)

Digitalni arhiv mrežnih publikacija (sustav DAMP ver. 1.0)

Miroslav Milinović
Sveučilište u Zagrebu
Sveučilišni računski centar - Srce
<Miroslav.Milinovic@srce.hr>

DAMP tečaj, 26.10.2004.

Sadržaj

- O projektu
- DAMP ver. 1.0.
 - dijelovi sustava
 - funkcionalni model sustava
 - razmjena podataka s Katalogom NSK
 - upravljanje pobiranjem
 - statusi u obradi
 - skladište podataka
 - pristup arhivi
- Pravci daljnjeg razvoja

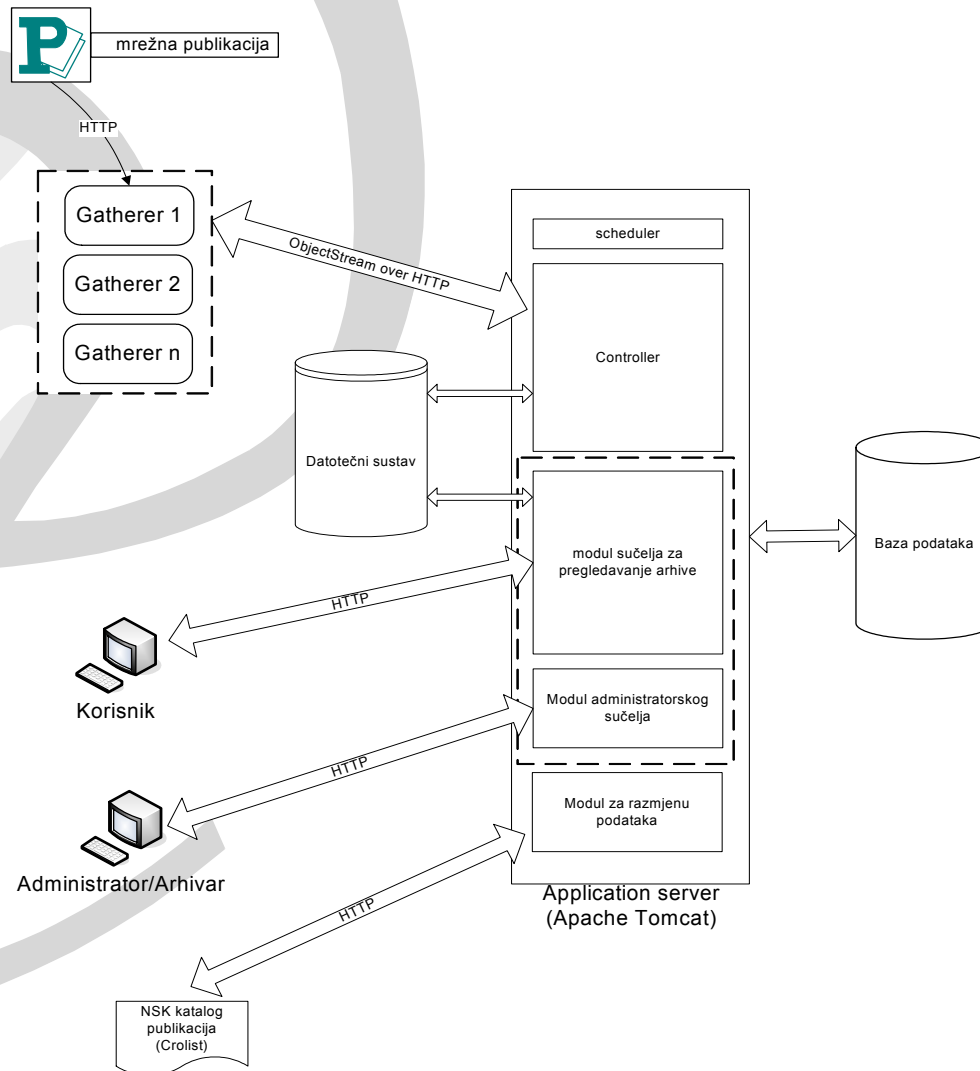
O projektu

- Projekt uspostave sustava za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija
 - zajednički projekt NSK i Srca
 - započet u studenom 2003. godine
 - cilj: uspostaviti sustav za preuzimanje i arhiviranje uz očuvanje autentičnosti sadržaja, oblika i funkcionalnosti mrežnih publikacija
- DAMP (**D**igitalni **A**rhiv **M**režnih **P**ublikacija) ver. 1.0.
 - dostupan putem <http://damp.srce.hr>
 - kontakt damp@srce.hr
 - modularan, proširiv, jednostavna uporaba za korisnika/arhivara
 - utemeljen na Open Souce programskoj podršci
 - razvijen na temelju iskustava Srca u području istraživanja Weba (projekt MWP)

Dijelovi sustava DAMP

- **Gatherer**
 - podsustav za pobiranje Web sjedišta (publikacije)
- **Controller**
 - podsustav za kontrolu pobiranja i arhiviranje rezultata pobiranja
- **Scheduler**
 - podsustava za raspoređivanje pobiranja
- **Skladište podataka**
 - čine ga baza podataka i posebno organizirani datotečni sustav
- **Web sučelje** za pristup arhivu i upravljanje sustavom
- **Modul za razmjenu podataka** s Katalogom NSK

Funkcionalni model



Razmjena podataka s Katalogom

- odvija se automatski (bez posredovanja administratora)
- DAMP je neovisan o Katalogu
- trenutno stanje:
 - razmjena podataka sa sustavom Crolist
 - DAMP kupi temeljne informacije o publikacijama, kandidatima za arhiviranje iz Crolista
 - DAMP prenosi informacije o arhivskim primjercima (URL adrese za pristup) u Crolist

Upravljanje pobiranjem

- administrator/arhivar mora obraditi svaku novu publikaciju:
 - odrediti frekvenciju (vrijeme i učestalost) pobiranja
 - odrediti parametre pobiranja
 - provjeriti ishod prvog (probnog pobiranja) i eventualno dodatno podesiti uvjete/parametre
- temeljem postavljenih uvjeta i parametara scheduler uvrštava konkretno pobiranje u red
- gatherer(i) izvršava(ju) pojedina pobiranja prema tom redu

Parametri pobiranja

- omogućuju administratoru/arhivaru detaljnije upravljanje pojedinim pobiranjem
- poželjno poznavanje Web tehnologija
- parametri:
 - **recursion_depth**
 - synonym
 - alternative_host
 - wanted_path_pattern
 - unwanted_path_pattern
 - remove_url_param
 - archive_extension
 - always_get_embedded_resource
 - get_external

Statusi u obradi

- Statusi koje ima pobiranje u redu (queue status)
- Statusi obavljenog pobiranja (uz arhivski primjerak)

Queue statusi

- **new** - novi u redu; još nije bilo pokušano pobiranje
- **waiting** - pobiranje se vratilo s greškom koja nije fatalna; bit će pokušano ponovno (postoji limit)
- **scheduled** - u obradi (pobiranje u tijeku)
- **wait for upload** - u obradi (u tijeku je zapisivanje rezultata pobiranja)
- **done** - privremeni status; ispravno obrađeno - briše se iz reda
- **error** - fatalna greška pri pobiranju ili potrošen broj pokušaja (vidi waiting)
- **upload error** - greška pri zapisavanju rezultata u arhivu
- **gatherer error** - greška gathrera
- **controller error** - greška controllera

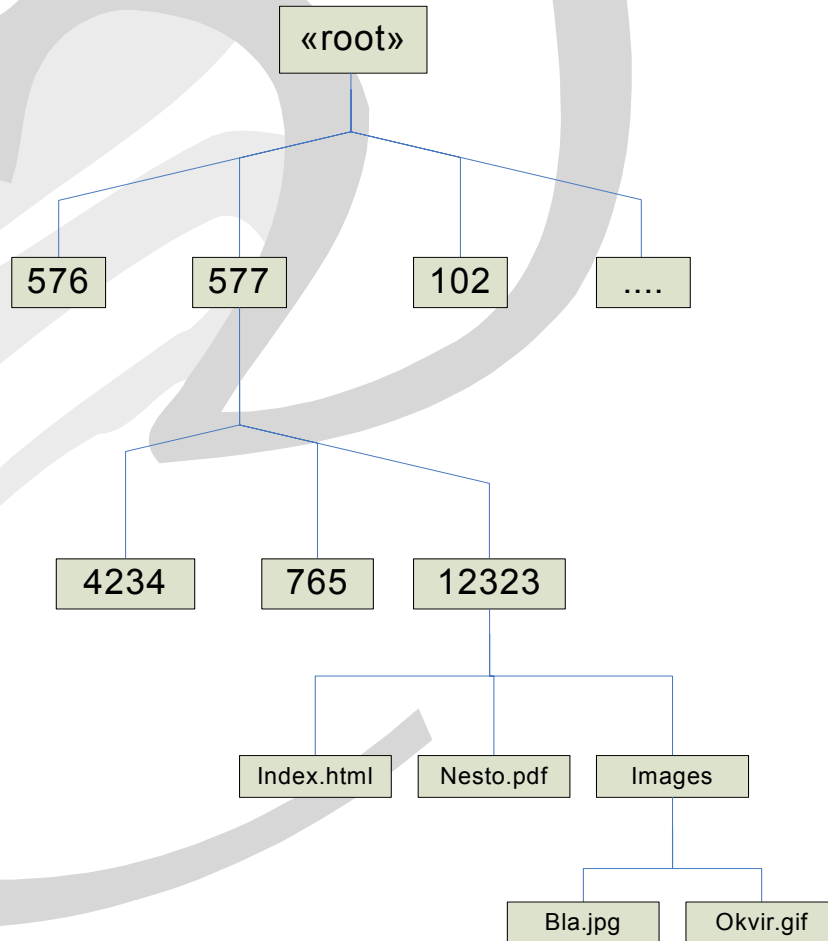
Statusi pobiranja

- **WAITING** - čeka početak obrade ili je u obradi
- **DONE** - pobiranje dovršeno bez grešaka
- **WARNINGS** - pobiranje dovršeno tako da barem jedan resurs nije uspješno pobran (HTTP status pri dohvatu > 400)
- **ERROR** - došlo je do problema s pristupom Web sjedištu tijekom pobiranja
- **FATAL ERROR** - greška u pobiranju

Skladište podataka

- baza podataka i posebno organiziran datotečni sustav
- u bazi podataka pamte se podaci o:
 - publikacijama (Web sjedištima)
 - parametri pobiranja
 - rezultati pobiranja
 - adresa arhivskog primjerka
 - ishod (status) pobiranja
 - metapodaci
 - distribucija tipova podataka za prikupljene datoteke

Datotečni sustav



početni dir.
datotečnog sustava

ID publikacije

ID pobiranja

početni dir. publikacije

Pristup arhivi

- Putem Web sučelja
- Za svako pobiranje pojedine publikacije omogućuje:
 - pregledavanje arhivskog primjerka
 - pregledavanje podataka o pobiranju
 - osnovni podaci o pobiranju (veličina, status, ...)
 - distribucija tipova prikupljenih datoteka
 - metapodaci
 - pregledavanje zapisa o pobiranju
 - izravni pristup direktoriju s rezultatom pobiranja

Pravci razvoja

- Nastavak suradnje NSK i Srca na poslovima:
 - izrade korisničkog sučelje (s odgovarajućim sustavom autentikacije i autorizacije) za javni pristup sustavu DAMP u svrhu pregledavanja arhivirane građe
 - pobiranja i arhiviranja mrežne građe kojoj je pristup zaštićen nekim mehanizmom autentikacije/autorizacije
 - izrade alata za prihvaćanje i arhiviranje arhiva nakladnika
 - izrade modula za složeniju dijagnostiku rada i statističku analizu sadržaja arhiva
 - održavanja i unapređenja (temeljnog dijela) sustava DAMP

Digitalni arhiv mrežnih publikacija (sustav DAMP ver. 1.0)

<http://damp.srce.hr>

damp@srce.hr