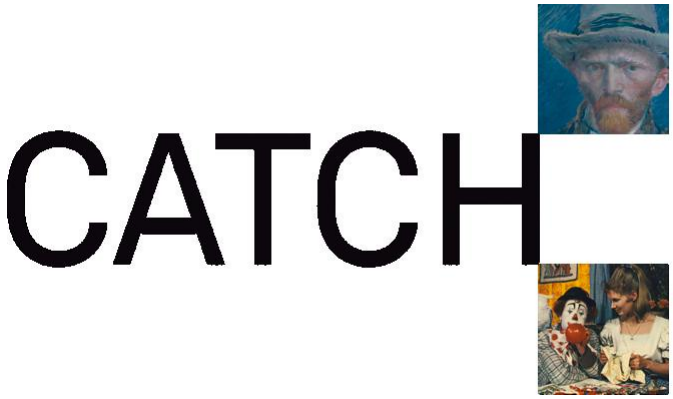




CONTINUOUS ACCESS TO CULTURAL HERITAGE PLUS



D1.1.1

Specificatie en implementatie van een REST web service

Hennie Brugman

1 Documenteigenschappen

Rapportage datum: 15/09/10
Rapportage periode: 01/06/09 tm 05/01/10
Versie: 1.0 ~~Concept~~ / definitief
Eventuele wijzigingen tov vorige versie? NVT

1.1 Document locatie

Bestandsadres:	http://www.catchplus.nl/wp-content/uploads/2010/09/D1.1.1-VAS-v1.doc

1.2 Historie

Versie	Datum	Veranderingen (concept/definitief)	Opdrachtgever	Auteur(s)
1.0	15/09/10	nvt	Marius Snyders	Hennie Brugman

1.3 Goedkeuring

Dit document heeft de volgende goedkeuringen nodig.

Naam	Rol	Opmerkingen	Datum
projectleider	verantwoordelijke		
projectbureau	advies		
Stuurgroep	akkoord		

1.4 Distributie

De volgende personen hebben een exemplaar ontvangen.

Naam	Functie	Datum
Marius Snyders	Projectleider CATCHPlus	15/09/10

Inhoud

1 Documenteigenschappen 3
1.1 Document locatie 3
1.2 Historie 3
1.3 Goedkeuring 3
1.4 Distributie 3
Management samenvatting 5
2 Bericht van oplevering 6
2.1 Paragraaf 1 ..... **Fout! Bladwijzer niet gedefinieerd.**
3 Gedetailleerde beschrijving van de deliverable 7
3.1 Paragraaf 1 ..... **Fout! Bladwijzer niet gedefinieerd.**
3.1.1 Subparagraaf 1 ..... **Fout! Bladwijzer niet gedefinieerd.**
4 Licentie 9
5 Demonstraties, toepassingen 10
6 Performance ..... **Fout! Bladwijzer niet gedefinieerd.**
7 Conclusie 11
Bijlagen ..... **Fout! Bladwijzer niet gedefinieerd.**

Management samenvatting

Door middel van de Vocabulary and Alignment Service (VAS) kunnen uiteenlopende terminologiebronnen, zoals in gebruik binnen erfgoedinstellingen, op een uniforme wijze centraal aan worden geboden voor gebruik door menselijke gebruikers en, vooral, computerprogramma's. Dit stelt CATCHPlus deelprojecten en andere tool-bouwers in staat hun software zodanig te bouwen, dat deze onafhankelijk is van welke specifieke terminologie-bronnen worden gebruikt. Bovendien biedt het aan aanbieders van terminologie-bronnen een eenvoudig middel om deze op verschillende manieren gestandaardiseerd, browse-baar en doorzoekbaar online aan te bieden (via een REST API, als Linked Open Data, als SPARQL endpoint). Deze bronnen dienen daartoe in SKOS formaat te worden aangeleverd, volgens door CATCHPlus opgestelde richtlijnen.

Deze deliverable betreft versie 1 van de Vocabulary and Alignment Service (VAS). Basis van de VAS is een installatie en configuratie van een Virtuoso RDF store. Via een in het kader van de VAS ontwikkelde procedure zijn in SKOS gerepresenteerde vocabulaires in deze RDF repository te laden of als geheel te updaten.

Deze versie van de VAS bevat een twaalftal vocabulaires, waarvan voor slechts twee al een open licentie is geregeld.

De RDF store is over het internet toegankelijk gemaakt door middel van een Java servlet, die de REST API implementeert. Functionaliteit van deze API omvat diverse zoekoperaties op concepten en vocabulaires, en het specificeren van de gewenste vorm van de geretourneerde resultaten.

Deze implementatie is, met enkele kanttekeningen, conform de specificaties. Hij wordt inmiddels in een tweetal relevante praktijkcases succesvol ingezet. Het conceptuele idee achter de VAS, de opzet van de huidige API en de implementatie zelf hebben inmiddels positieve respons opgeroepen bij thesaurus-aanbieders en bij relevante bedrijven. VAS heeft daarmee goede kansen breder in de nederlandse erfgoedsector ingezet te worden, en kan beschouwd worden als een succesvolle eerste stap naar valorisatie.

2 Bericht van oplevering

2.1 Wat

Versie 1 van de Vocabulary and Alignment service (VAS). Deze VAS bestaat uit installatie en configuratie van een tweetal Virtuoso RDF stores, één ten behoeve van testen en software-ontwikkeling, en één voor daadwerkelijk gebruik. De content van beide RDF stores bestaat uit een twaalftal naar SKOS geconverteerde thesauri. Verder is in eigen beheer een Java servlet ontwikkeld, die enerzijds een REST API implementeert, en anderzijds toegang tot de RDF store middels JDBC regelt. 'Bovenop' ieder van de twee RDF stores is een instantie van die servlet geïnstalleerd.
Daarnaast zijn een API specificatie (online), technische documentatie plus installatie-handleiding, en source code (in een online source code repository) opgeleverd. Hosting van alle onderdelen is geregeld bij een externe partij (Tuxic), waarbij Beeld en Geluid de opdrachtgever is.

2.2 Wie, wanneer en waar

De oplevering heeft op 5 januari 2010 bij Beeld en Geluid plaatsgevonden. Daarbij waren aanwezig: Hennie Brugman (technisch coördinator CATCHPlus) en Lourens van der Meij (software ontwikkelaar, ingehuurd van de Vrije Universiteit Amsterdam)

2.3 Mate van succes

De requirements die bij de aanvang van de werkzaamheden zijn gespecificeerd voor deze versie van de VAS zijn grotendeels geïmplementeerd. Er zijn een aantal nog niet geïmplementeerde features en bekende bugs, zie daarvoor hoofdstuk 3.

De VAS wordt momenteel in de praktijk gebruikt via een algemene browse- en zoek-webapplicatie (D1.1.9) en in een samenwerkingsverband tussen Beeld en Geluid en het Nationaal Archief. Daarvoor heeft de fa Pictura toegang tot de VAS ingebouwd in hun Memorix collectie-beheer-systeem.

Door de positieve ervaringen met de VAS benadering, API en implementatie is inmiddels brede belangstelling bij bedrijven en erfgoedinstellingen ontstaan voor participatie in de ontwikkeling van versie 2

3 Gedetailleerde beschrijving van de deliverable

Het basisidee van de vocabulary repository is, om uiteenlopende terminologiebronnen, zoals in gebruik binnen erfgoedinstellingen, op een uniforme wijze centraal aan te bieden voor gebruik door menselijke gebruikers en, vooral, computerprogramma's. Dit stelt deelprojecten en andere tool-bouwers in staat hun software zodanig te bouwen, dat deze onafhankelijk is van welke specifieke terminologie-bron ook wordt gebruikt. Bovendien biedt het aan aanbieders van terminologie-bronnen een eenvoudig middel om deze op verschillende manieren gestandaardiseerd, browse-baar en doorzoekbaar online aan te bieden (via een REST API, als Linked Open Data, als SPARQL endpoint). Deze bronnen dienen daartoe in SKOS formaat te worden aangeleverd, volgens door CATCHPlus opgestelde richtlijnen.

Deze deliverable betreft versie 1 van de Vocabulary and Alignment Service (VAS). Basis van de VAS is een installatie en configuratie van een Virtuoso RDF store. Via een in het kader van de VAS ontwikkelde procedure zijn in SKOS gerepresenteerde vocabulaires in deze RDF repository te laden of als geheel te updaten.

Deze versie van de VAS bevat een twaalftal vocabulaires, waarvan voor slechts twee al een open licentie is geregeld.

De RDF store is over het internet toegankelijk gemaakt door middel van een Java servlet, die de REST API implementeert. Functionaliteit van deze API omvat diverse zoekoperaties op concepten en vocabulaires, en het specificeren van de gewenste vorm van de geretourneerde resultaten.

3.1 Onderdelen van de deliverable

Diverse onderdelen van de deliverable zijn online te vinden (stand 15 september 2010):

Installatie- en configuratie-procedure: http://catchplus.tuxic.nl/project/vas/pdf/vas.pdf

Uitvoerige documentatie:
http://www.catchplus.nl/wp-content/uploads/2010/09/documentation-vas-v1.doc

Source code is te vinden in de volgende Subversion repository:
svn+ssh://catchplus@catchplus.tuxic.nl/usr/local/catchplus/svnroot/vocabularyservice

API specificatie: http://catchplus.tuxic.nl/catchplus/serviceapi/1/serviceapi.html

De service is online beschikbaar op: http://catchplus.tuxic.nl:18080/vas/api/

3.2 Bugs en omissies

Bekende bugs/omissies ten opzichte van de oorspronkelijke specificaties:

- De API ondersteunt nog geen filteren op taal
- Html output format voor concepten, conceptschemes en conceptschemegroups ontbreekt
- Rdf en html output format voor alignments ontbreekt
- De API zou hier en daar resultaten moeten ontdubbelen
- Upload en download via de REST API ontbreekt nog

Deze bugs een ommissies worden in versie 2 van de VAS geadresseerd (D1.1.4).

Verder kent deze versie van de VAS enige beperkingen:

- Er is niet uitgebreid getest op schaalbaarheid. Ervaring met RDF stores als Virtuoso leert dat schaalbaarheid een probleem kan blijken te zijn. Daarom zal hier voor versie 2 in de specificatie-fase al aandacht aan worden besteed.
- Doorontwikkeling van de huidige oplossing vergt specifieke expertise, waardoor overdracht van de huidige implementatie aan een andere partij tijdrovend kan blijken te zijn.

3.3 Toekomstplannen

Concreet moet worden gewerkt aan een geavanceerder distributiemodel, performance issues (voor zover die er blijken te zijn), en meer high level API calls, om waar nodig snel veel informatie ter plaatse te kunnen krijgen. Verder zal ieder concept voorzien worden van een persistente identifier en/of stabiele http URL, en zal de repository content worden gepubliceerd als Linked Open Data. Daarnaast zullen waar mogelijk nieuwe thesauri worden toegevoegd. Tenslotte zal per thesaurus de licentieproblematiek worden opgepakt. Omdat niet iedere thesaurus met een open licentie beschikbaar zal komen, en voor operaties die de inhoud van de repository veranderen zal gebruikers-authenticatie worden geregeld op het niveau van de API.

Voor de implementatie zal worden gestreefd naar samenwerking in een consortium met een drie- of viertal bedrijven (Adlib, Pictura, Trezorix, mogelijk CQ2). De intentie daartoe is door al deze partijen inmiddels uitgesproken.

4 Licentie

Voor de VAS Java code geldt de GNU Public License versie 3.0.
Van Virtuoso is de open source versie gebruikt. Hiervoor geldt GPL 2.0

Voor de inhoud van de VAS repository is alleen voor de GTAA van Beeld en Geluid inmiddels een expliciete licentie geregeld: Open Database License (OdbL).

5 Demonstraties, toepassingen

Deze versie van de Vocabulary and Alignment Service kent momenteel twee toepassingen.

1. In het kader van deliverables D 1.1.2 en D 1.1.9 is een generieke browse- en zoekapplicatie ontwikkeld, die gebruikers in staat stelt de repository-inhoud te inspecteren. Deze web-applicatie draait autonoom, op een andere web server dan de VAS. Momenteel is een instantie van de browse- en zoek-applicatie te vinden op: http://vocrep.q42.net/index.html
2. In het kader van een samenwerking tussen Beeld en Geluid en het Nationaal Archief heeft de firma Pictura support voor de Beeld en Geluid thesaurus aan haar Memorix collectie-beheersysteem toegevoegd op basis van aanroepen van de CATCHPlus VAS.

Bij beide toepassingen is tot nu toe de stabiliteit, functionaliteit en performance van de VAS afdoende gebleken, met inachtneming van de onder hoofdstuk 3.2 genoemde omissies en bugs.

6 Conclusie

Deze deliverable is succesvol opgeleverd, wordt inmiddels in een tweetal praktijkcases gebruikt, en gaat mogelijk de basis vormen van een binnen het Nederlands digitaal erfgoed breed gedragen dienst.