



Paul Doorenbosch
STITCHplus - gebruikerscommissie

CATCH 1^e ronde (2005-2009)

Onderzoekslijnen

1. Semantische interoperabiliteit

2. Kennisverrijking met behulp van automatische analyse

3. Personalisatie in presentatie

STITCH (2005-2009)

Koninklijke Bibliotheek – Vrije Universiteit

Uitgangssituatie:

- Diverse catalogi
- Zo'n 25 ontsluitingssystemen (thesauri, begrippenlijsten, classificaties)
- Eén database (GGC (OCLC))

Vraagstelling:

Zijn ontsluitingssystemen geautomatiseerd aan elkaar te linken met behulp van semantisch web technieken, zodat afzonderlijke (deel)collecties semantisch integraal doorzoekbaar worden?

Methoden:

- Lexicaal (woordvergelijking)
- Instance based (gemeenschappelijke titelrecords)

STITCH (2005-2009)

Koninklijke Bibliotheek – Vrije Universiteit

Wetenschappelijk resultaat

- Methodieken voor conceptalignment
- Evaluatiemethodieken

Praktisch resultaat

- Algoritmes voor conceptalignment
- Trefwoordsuggestietool
- Prototype vocabulairerepository [integrator ism andere partijen]

STITCHplus

Koninklijke Bibliotheek

Doelen van STITCHplus, zoals in projectaanvraag

- Doorontwikkelen van algoritmes voor conceptalignment
- Implementatieversie van trefwoordsuggestietool
- Bouw vocabulairerepository

STITCH plus

Koninklijke Bibliotheek

Specifieke resultaten:

- 5-tal naar SKOS geconverteerde vocabulaires
- Lokale triple store
- Harvestingmechanisme tbv vocabulaire data
- Pijplijn module voor updaten centrale dienst
- Inventarisatie en doorontwikkeling alignment algoritmes
- Trefwoordsuggestietool
- Web GUI trefwoordtool
- Draft businessplan
- Documentatie, open source code

STITCH plus

Koninklijke Bibliotheek

Veranderingen in KB-beleid (2009-2010)

- Geen prioriteit voor onderwerpsontsluiting
- Geen prioriteit voor opschonen database (ontsluitingsgegevens)
- Stopzetten GOO

Overige ontwikkelingen

- Ontwikkelde algoritmes niet overdraagbaar
- Vocabulaire repository -> gemeenschappelijke dienst

STITCH plus

Koninklijke Bibliotheek

Specifieke resultaten:

- 5-tal naar SKOS geconverteerde vocabulaires
- Lokale triple store
- Harvestingmechanisme tbv vocabulaire data
- Pijplijn module voor updaten centrale dienst
- ~~- Inventarisatie en doorontwikkeling alignment algoritmes~~
- ~~- Trefwoordsuggestietool~~
- ~~- Web GUI trefwoordtool~~
- Draft businessplan
- Documentatie, open source code

STITCH plus

Koninklijke Bibliotheek

Afgesproken resultaten:

- 5-tal naar SKOS geconverteerde vocabulaires
- Lokale triple store
- Harvestingmechanisme tbv vocabulaire data
- Pijplijn module voor updaten centrale dienst
- Draft businessplan
- Documentatie, open source code

Toegevoegd:

- Geautomatiseerde matching Persoonsnamenthesaurus met VIAF
- Matching Persoonsnamen en onderwerpenthesauri aan Wikipedia
- Crowdsourcingtool om matchin te valideren
- Textanalyzer om named entities te herkennen en deze koppelen aan concepte
- Eventanalyse voor full text, met eventuele koppeling aan concepten

STITCH plus

Koninklijke Bibliotheek

- 5-tal naar SKOS geconverteerde vocabulaires ✓
- Lokale triple store ✓
- Harvestingmechanisme tbv vocabulaire data ✓
- Pijplijn module voor updaten centrale dienst ✓
- Draft businessplan **in ontwikkeling**
- Documentatie, open source code ✓

- Geautomatiseerde matching persoonsnamenthesaurus met VIAF ✓
- Matching persoonsnamen- en onderwerpenthesaurus aan Wikipedia ✓
- Crowdsourcingtool om matching te valideren ✓
- Textanalyzer om named entities te herkennen en deze koppelen aan concepten ✓
- Eventanalyse voor full text, met eventuele koppeling aan concepten ✓

STITCH plus

Laatste stappen

- Licentie thesauri
- Draft businessplan
- Verrijking persoonsnamenthesaurus
- Koppeling ANP-collectie aan GTAA (via OpenSKOS [vocabulairebank]) (?)

Vragen?

Demonstraties door René van der Ark

