Effective retrieval with social tags

Mettina Veenstra mettina.veenstra@telin.nl
Telematica Instituut

Hilversum, March 7th 2008

Telematica Instituut

Topics

- Context
- Research questions tagging experiment
- Results tagging experiment
- Conclusions

Kennisnet Ict op school

SURF NET

Teleblik Voortgezet onderwijs

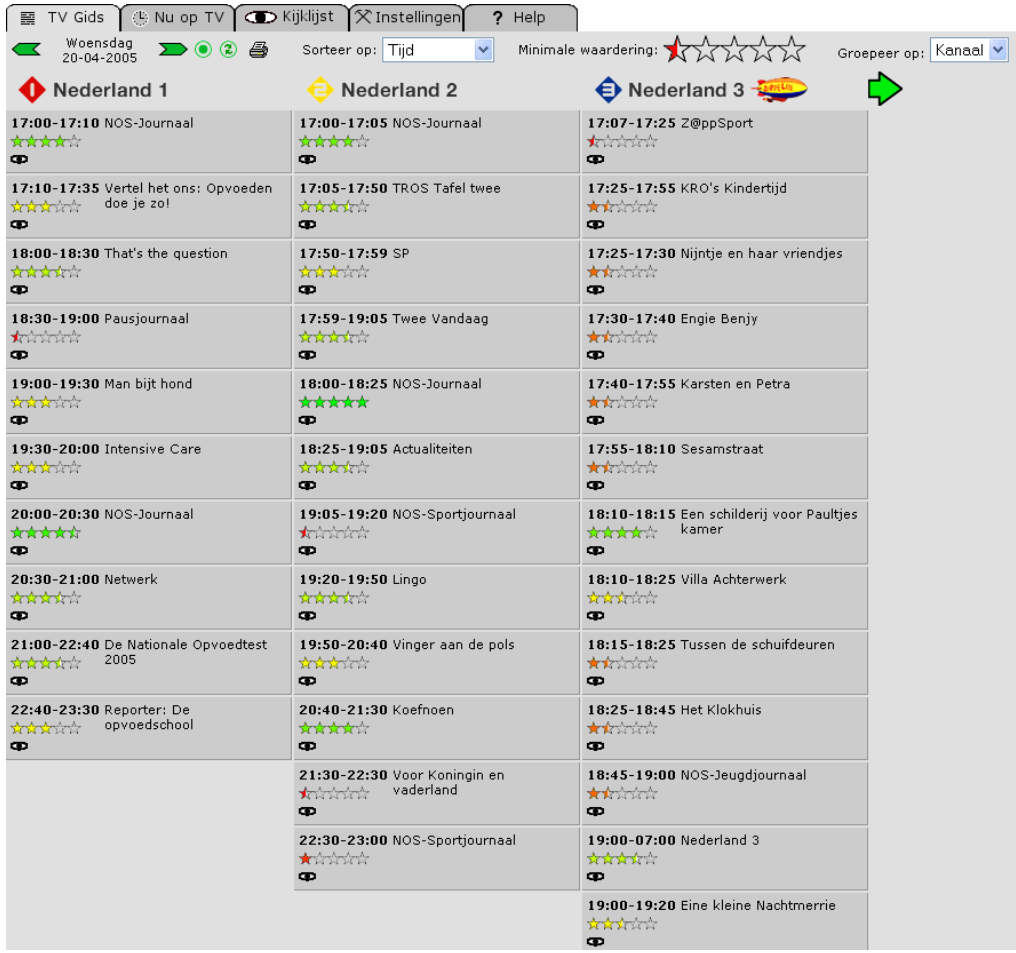
Telematica Instituut

Context (research)

- How can people interact with large collections of video in an efficient way?
  - Find a specific video they are looking for
  - Receive recommendations for interesting videos
  - Browse to or find interesting fragments within a video
  - Browse from one interesting fragment to comparable other fragments within or outside the current video
  - >>> emphasis on recommendation and video browsing
- Good quality metadata and profiles are necessary
  - >>> Emphasis on text analyses and social tagging

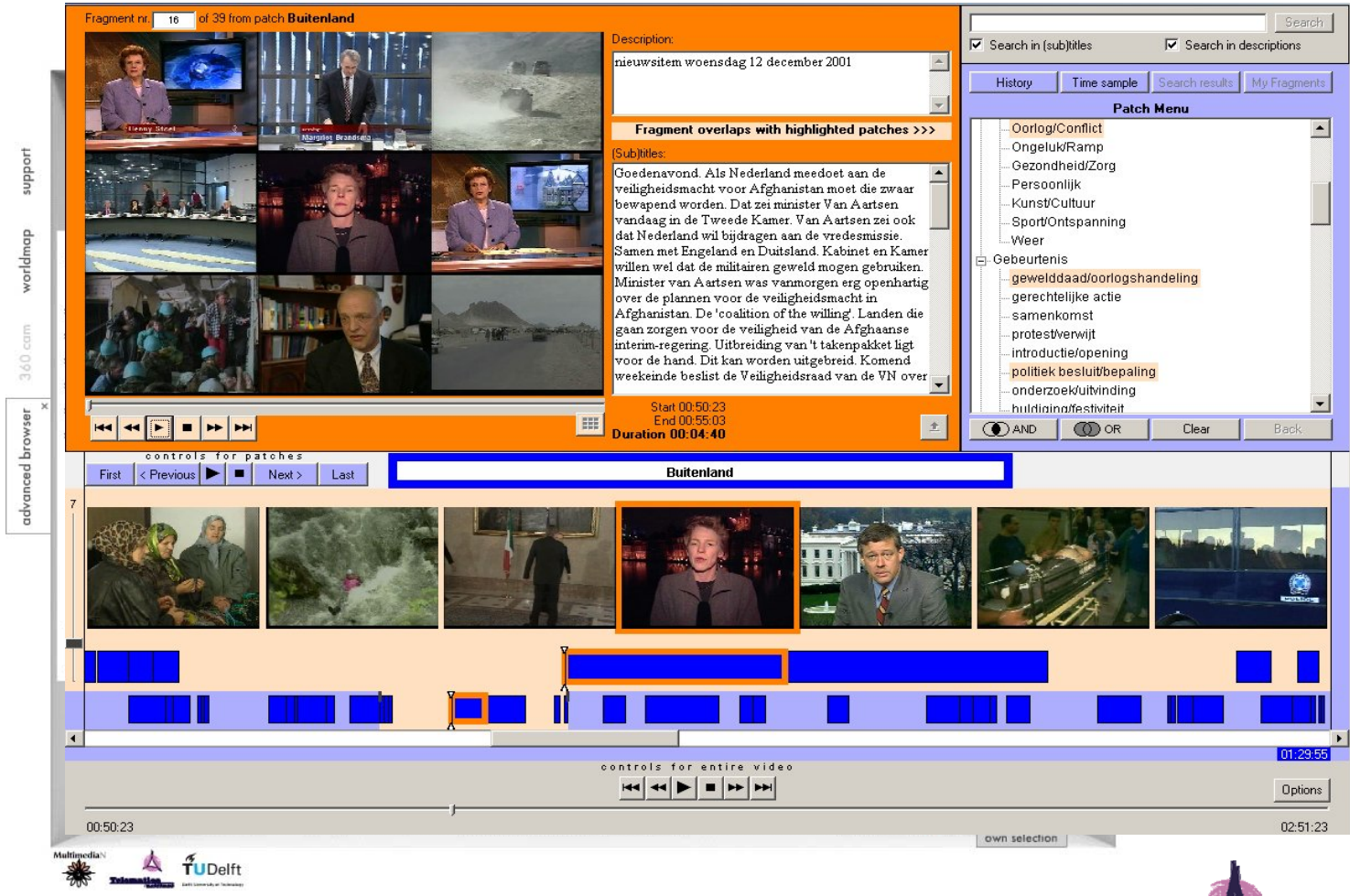
Telematica Instituut

Context: recommendation



Telematica Instituut

Context: video browsing



Telematica Instituut

Context: metadata > content analysis

Missie Afghanistan uiterst onzeker

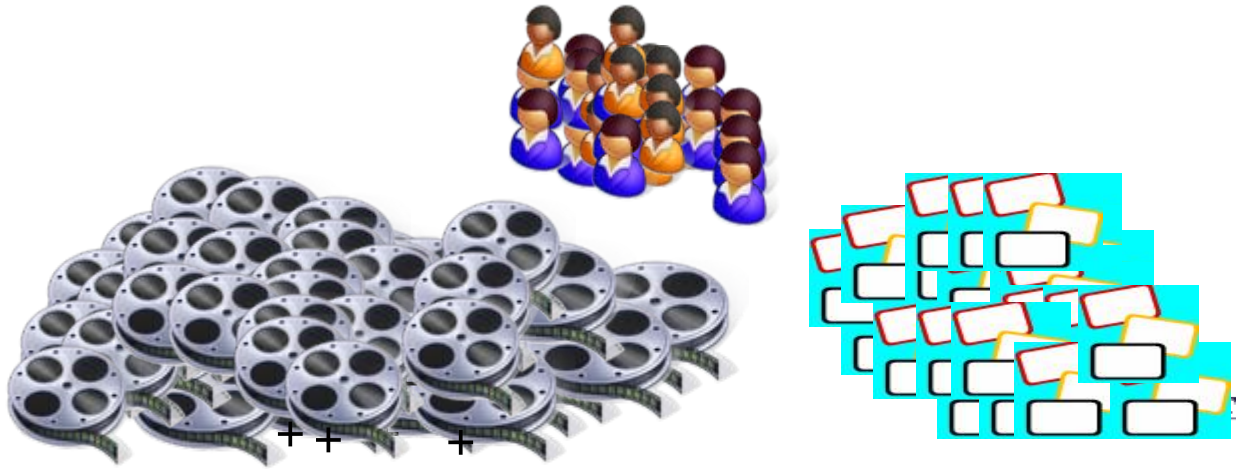
Steeds meer partijen beginnen te twijfelen aan de voorgenomen missie van 1100 Nederlandse soldaten naar Afghanistan. Morgen komen er twee hoge functionarissen van het Pentagon en het State Department naar Den Haag voor overleg met Nederlandse topambtenaren. Vrijdag hakt het kabinet zo goed als zeker de knoop door. Het lijkt een ware worsteling te worden.

Title
Summary
Genre
Subject
Person names
Names
Locations
Maker name
Speaker

Context: more metadata > social tagging

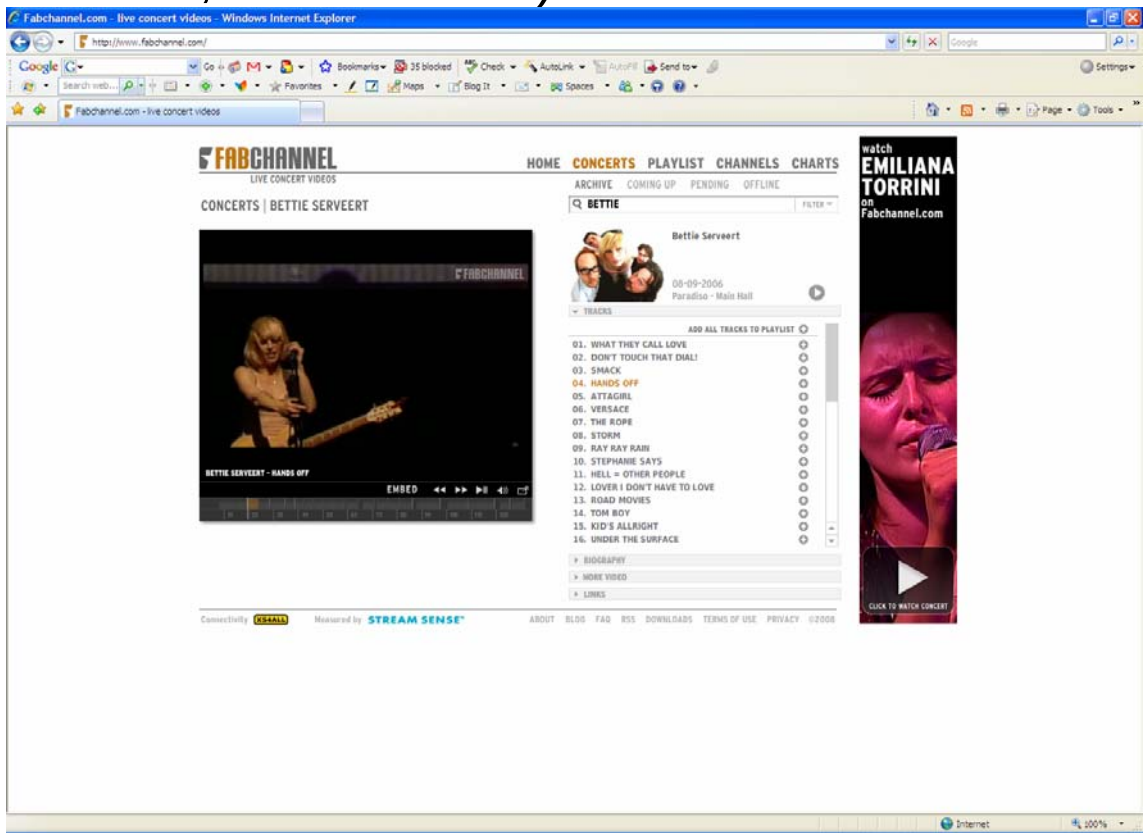
- More and better metadata?
  - Do your peers have more comparable vocabularies than professionals
  - How good are social tags in comparison with:
    - professional metadata?
    - automatically generated metadata?

this presentation



Context: summary

- What is the main motivation behind this research?
  - Give users more possibilities than just retrieving, starting and stopping a video (YouTube, Google video, broadcasters)



Telematica Instituut

MultimediaN ViTa research 2007/2008

- How helpful are tag suggestions for taggers? Assumption: choosing is easier than “being creative”, thus:
  - Users could be quicker at completing their tasks
  - It could result in more tags per video
- How suitable are tags for finding relevant results? More precisely: How is the quality of tags in comparison with professional metadata or automatically generated metadata?
- Public/cultural/educational value, Commercial value, Social and community value
- Critical success factor: Cataloguing (more and possibly very detailed information on a segment level)

Telematica Instituut

1. Tagging phase

Video content: 115 videos selected for educational purposes (culture), 60-150 seconds, already enriched with metadata.

194 students, 3 experimental conditions, each person should tag 15 videos.

	Social Tag Suggestions	Metadata Tag Suggestions
Basic Tagger	⊘	⊘
Social Tagger	✓	⊘
Lazy Tagger	✓	✓

Telematica Instituut

Welkom bij Vita (Development Versie) - Windows Internet Explorer

http://client342.lab.telin.nl:8081/vitafrontend-0.1-SNAPSHOT/nl.telin.vita.Main/Main.html;jsessionid=E6B37600EA416C4934F0AA5AB3

Links » File Edit View Favorites Tools Help

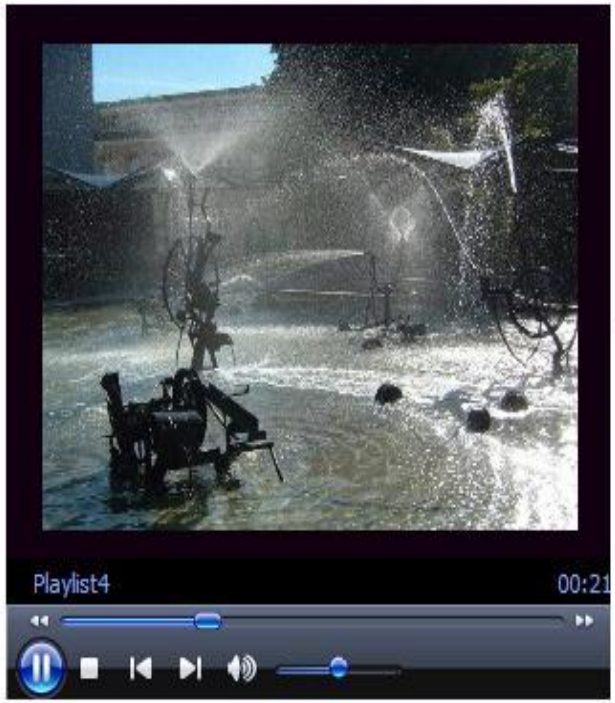
Welkom bij Vita (Development Versie)

Page Tools

Welkom BasicSocialSearcher !

Help

Video - Bewegende kunst is mooier wanneer het groot is gemaakt



Titel : Bewegende kunst is mooier wanneer het groot is gemaakt

Duur : 00:01:18 (hh:mm:ss)

Beschrijving : Het is dan spannender om naar te kijken. Wanneer er iets klein draait is dat niet zo veel aan. Maar een grote bewegende installatie is veel boeiender om naar te kijken. Het duurt wel langer om een grote installatie te maken. Soms duurt het een jaar voordat het kunstwerk af is.

Mijn tags:

Tag	Type

Tagsuggesties:

- kunst
- kunstwerk
- kunstwerken
- technologie
- muziek
- bewegende kunst
- beweging
- kunstenaars
- bewegend
- kunstvoorwerpen

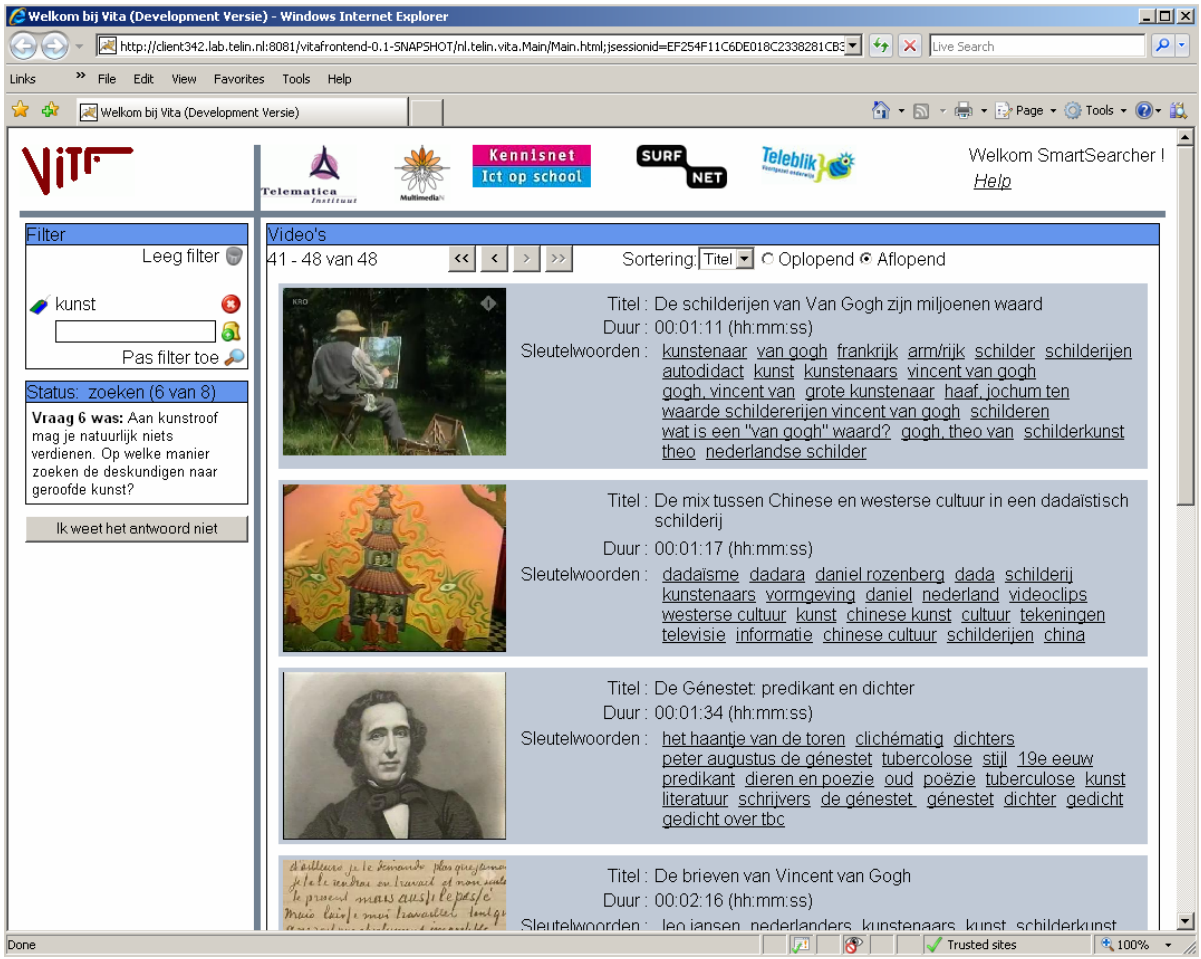
2. Search phase

	Professionele Metadata	User Tags	Smart Metadata	Smart tags
BasicSearcher	✓	✗	✗	✗
SocialSearcher	✗	✓	✗	✗
BasicSocialSearcher	✓	✓	✗	✗
DocumentSearcher	✗	✗	✓	✗
SmartSearcher	✗	✗	✗	✓

153 students, 5 experimental conditions, 8 questions to answer.

Telematica Instituut

Search conditions



smart

Telematica Instituut

Results of the tagging phase 1/1

Origin adopted tag suggestions	SocialTagger	LazyTagger
Originating from metadata:	44%	57%
- Keywords	16%	18%
- Derived from description	28%	39%
Originating from other taggers:	56%	43%

- Many social tags adopted -> quality. (SocialTaggers have the lowest number of unique tags)
- Tags of LazyTaggers influence the SocialTagger, who thus *indirectly* also adopts tags based on metadata
- Taggers that receive suggestions are in general not faster than others, nor do they create more tags

- Despite of the suggestions the SocialTagger and the LazyTagger create a lot of tags themselves:
  - 57% adopted vs. 43 created (not in this table)

Telematica Instituut

Results of the search phase 1/2

	Answered	Correctly answered
Role	*Mean*	*Mean*
Basic*	6.4	4.5
Social*+	6.7	5.5
BasicSocial*+	6.7	5.0
Document	5.3	3.5
Smart+	6.2	4.8
Total	**6.2**	**4.6**

* “answered”: difference with DocumentSearcher is statistically significant
+ “correct”: difference with DocumentSearcher is statistically significant

Trends:

- The “Social” groups – with tags – perform best
- They also needed the least number navigation steps per successful answer
- Document searchers do not do too well

Telematica Instituut

Results of the search phase 2/2

8616 search terms:

- 96% entered in the search field
  - 58% from questions
  - 13% correspond to tags
  - 4% correspond to professional or automatically generated metadata
  - > This gives an indication that tags fit in with the users' vocabulary
  - 25% invented by the user
- 4% clicked on in the interface
- > clicking on terms in the interface doesn't seem to be a natural thing to do (Google generation?)

Telematica Instituut

Conclusions

- Social tags better than professional metadata?
  - Well, at least not worse in this case and maybe even better?
  - The terminology seems to correspond more to the users own terminology
  - Pollution: The quality of the tags wasn't a problem in our case
    - But this was not a natural setting
    - With large numbers of users "bad" tags can be filtered out
- The amounts of content seem to make the need to seriously consider social tagging (and automatic metadata extraction) more urgent
- Future research:
  - How to motivate people to tag? (paying, gaming, helping to improve own content organisation). Social and community value
  - How useful are tags for personal profiles? The importance of who tagged what.

Telematica Instituut

Questions?

Mettina Veenstra: mettina.veenstra@telin.nl

Telematica Instituut