Flexible Services

CrossMedia
D1.5 Knowledge-based services embedded in editorial systems

Document Name	D1.5 Knowledge-based services embedded in editorial systems
Project/WP Title:	CrossMedia
Document Type, Security	Work in progress

Document Title:	D1.5 Knowledge-based services embedded in editorial systems
Agreed date of delivery	30.11.2009
Actual date of delivery	30.11.2009
Editor	Asta Bäck
Version and	version 1.0
Date Last Change	30.11.2009
File:	FS_CrossMedia_D1.5_v_1.0.pdf

Participants	Name	e-mail
VTT	Bäck, Asta	Asta.Back@vtt.fi,
	Vainikainen, Sari	Sari.Vainikainen@vtt.fi
Profium	Saarela, Janne	Janne.Saarela@profium.com

Executive Summary

This deliverable summarises the results of the work that was carried out in the Task 1.5 “Enhance technologies” of the CrossMedia project. The task concentrated on developing technologies, which will be tested with media companies and users in Spring 2010 in the cases.

This document consists of two parts. The first part deals with creating and utilising geolocation information in media companies’ operations. Geolocation information will be widely available both on media creator and consumer sides. In addition to exact coordinates, other types of GIS primitives such as lineStrings and polygons should be supported in the future editorial systems.

The second part deals with the concept of a service-independent user profiling service and presents the technologies needed for creating semantic profiles of users’ interests, and using the profiles for personalisation.

We have described the current implementation of the service-independent user profiling service in this document. The service was implemented by utilising VTT’s existing background applications as the starting point. The profile service was developed further into an independent service. Updated features include user registration and the development of a database for storing profile data, and improved analysis of users’ tags, which now covers also Finnish and Swedish in addition to the earlier English. For adding semantics to interests, various semantic knowledge bases are used. VTT’s background application used only WordNet, but now also KOKO, DBpedia and Freebase are used.

APML (Attention Profiling Markup Language) with some extensions was selected as the format for sharing profile data between services. Other options for representing profiles are presented and their features discussed.

1. Tables and Figures

List of figures

List of Acronyms and Abbreviations

APML	Attention Profiling Markup Language The vocabulary for exchanging attention data between applications
DBpedia	DBpedia (http://dbpedia.org/) is a community effort to extract structured information from Wikipedia and to make this information available on the Web.
FOAF	The Friend of a Friend ontology
Freebase	Freebase (http://www.freebase.com/) is an open database built by the community. It contains structured information on many popular topics, including movies, music, people and locations.
GeoNames	The GeoNames (http://www.geonames.org/) is a geographical database
GIS	Geographic Information System
GML	Geography Markup Language
KML	Keyhole Markup Language
KOKO	ONKI ontology library service (http://www.yso.fi/) contains ontology such as KOKO Ontology which is a collection of Finnish core ontology.
MOAT	Meaning of A Tag Ontology
OGC	Open Geospatial Consortium
RDF	Resource Description Framework
SCOT	Social Semantic Cloud of Tags Ontology
SIOC	Semantically Interlinked online-communities ontology
SKOS	Simple Knowledge Organization System ontology
SPARQL	SPARQL Protocol and RDF Query Language
WFS	Web Feature Service
WMS	Web Map Service
WordNet	WordNet is a large lexical database of English

2. Introduction

This report, Knowledge-based services embedded in editorial systems, looks at opportunities to add new, intelligent features into media services. Knowledge in this case refers to semantic metadata, and knowledge-based services are services that utilise this knowledge. Semantic metadata that builds on the Linked data principle opens the way to utilise the knowledge gathered into various knowledge bases available on the internet.

This report focuses on two services:

- Adding location related knowledge into content, and
- Creating and using user's interest related profiles for personalisation.

The main focus in the document is on issues dealing with service-independent user profiles: how such profiles can be created and how they can be connected to media content and consumption.

3. Geolocation

by Janne Saarela, Profium

Media companies have increasing interest to contextualize their products and services both for the end-users and for the advertisers as it adds value for both of those parties. This contextualization can take many factors into account. This chapter focuses on making use of geographic information in media companies operations. One should stress that geolocation is not an island of its own but it is one of the factors that can be taken into account when creating personalized services and/or targeted advertising.

3.1. Creating geolocation information

As with other types of metadata some burden from its creation can be relieved by automatic measures. Such measures include for example digital cameras and mobile phones with embedded GPS receiver and software ability to include GPS location (point) within images or other type of content object being created.

Importing the geolocation for public or private sources is another automated option for creation geolocation information. For example street networks and or property limits can be retrieved with batch or on-demand queries from other IT systems. Here the types of primitives could be encoded as lineStrings i.e. a set of points which approximate for example a street or as a polygon which approximates an area such as city limits.

Manual addition of geolocation information asks for map interfaces where the user can then with the help of a mouse click and determine a point, lineString or a polygon that needs to be created as additional knowledge for a content object.

3.2. Making use of geolocation information

Once geolocation information has been encoded in computer understandable form, we can let algorithms determine if some content is of relevance to a user or if an ad should be associated with the content. Topological functions such as inside and intersects can be used to determine if, for example, a story about an upcoming railroad strike (i.e. temporal dimension indicates the story is still valid and not historical) with lineString information can be calculated to be within the boundaries of the city where registered user is entering an online service. Equally, an advertiser for locally available services such as a restaurant can ask its advertisements to be shown if a story about an event is within 2 kilometers. This example requires support for spatial functions capable of calculating distance between different GIS primitives.

3.3. Applicable technologies for geolocation

Open Geospatial Consortium (OGC) is an industry initiative that develops standards for geospatial and location based services. OGC promotes the use of these standards as a vendor neutral way of encoding GIS information.

Geography Markup Language (GML) allows encoding geographic information within XML encoded documents. Web Map Service (WMS) allows exposing maps to WMS capable client software and Web Feature Service (WFS) allows querying and exposing various features on such maps.

Keyhole Markup Language (KML) was originally submitted to OGC by Google and is now supported in their Earth service in OGC processed version 2.2.

These technologies form the core of geolocation capability supported across a wide range of GIS vendors. In order to enable smoother integration with other types of knowledge used by media companies there is work in progress to unify GIS representation and processing using Semantic Web technologies and RDF and SPARQL in particular.

3.4. Geolocation in the context of editorial systems

State-of-the-art in editorial systems has until today exhibited support for encoding city/country geolocation information with content and then, with the use of approximation, the users have been able to visualize this information on a map. Editorial systems should today be able not only to associate the content being

prepared with an exact coordinate point but associate it with other types of GIS primitives such as lineStrings and polygons as well. We have given motivation for their use in the previous discussion.

Finally, in the domain of news, the notion of an impact area asks for a parallel knowledge extension for content: something where the story will have an impact as a function of time. The typical sample would be an oil leak in a river where the actual story is associated with a point but the impact area can either be a lineString (approximation of the downstream river) or a polygon (approximation of the riverbanks downstream river).

4. Service-independent user profiles

by Asta Bäck and Sari Vainikainen, VTT

4.1. Motivation for a service independent user profile

Personalisation is the technology that holds the promise to serve users better in different kind of information services. When the content can be selected and prioritised based on users' interests and intentions in a particular situation, user experience is improved. There are several challenges relating to personalisation. The main challenges are how to create and have access to a relevant and descriptive user profile, how to understand the user's context and intentions correctly at a given time, and how to combine these pieces of information in order to offer the best possible content or service to the user. Of course, it is also necessary that the media company has content that is described with semantic metadata and that there is content that matches the identified needs and interests.

Personalisation can be used in media companies relating to the actual media content, and advertisements, and any other services that a media company has to offer to its customers. If profile information is available and users give the necessary permissions, also recommendations of people could be made.

Instead of each company trying to put together user profiles from scratch, we vision an independent service for creating portable, service-independent user profiles. This service would be the place, where each user creates one or several profiles, which she can then choose to share with different media services in order to get a personalised media experience.

It is reasonable to assume that if one profile can be used in many services, users are more motivated to create such a profile. Potentially, users could also be given a chance to use this profile in their own sites, such blogs and social networking sites, if a widget is offered to display selected parts of the profile. Another feature that may encourage people to create a profile is to let people decide

freely whether the profile is an anonymous one, or connected to their real personality.

4.2. About user profiles

A User profile may describe any and many aspects of a person, and it depends on the application area, which aspects are relevant and important. Traditionally, In connection to media and advertising, demographic information describing age, gender, marital status and the number and age of children are typically used. Many interest areas can be inferred based on this information.

Interests and hobbies are another important part of a person's profile. In traditional print publishing world, it was important to estimate the amount of interest to certain topics, and then, if the amount was estimated to be big enough, an article or publication was offered relating to it. Now, with electronic publishing, users' interest's can be served in much shorter cycle, in a number of channels, and at small granular level.

Currently, search engine providers and advertising networks have a lot of information about people's interests. Advertising networks typically tract the behaviour of a user recording what kind of sites a user visit and based on that infer areas of interest. In the behavioural model, little or no attention is given to traditional demographic data, which may be difficult to obtain in this context. Rather, the assumption is that, if a person visits a site specialising in a certain topic, advertisements relating to that topic are relevant to this person. Large sites, like Google, Yahoo! and Facebook may gather a lot of information relating to a specific user, if and when the user has logged into their systems while making searches and using other features. This information is accessible only to these companies, and gives them a major advantage in developing their services.

4.3. Sources of profile information

A user profile may contain very different types of information, and there are different sources and opportunities for gathering this information. In the following list, we give some examples and ideas as what kind of different sources and ways there are to gather the information.

- In many social media services, users give explicit information of themselves on the profile page. Profile pages often have fields for different types of information, some of which is demographic, such as gender, age, and language, some relate to interests. Interests may typically be described with free text, but often users are also asked to tell about their favourite movies, books and artists. Wigdets, such a countries visited, are popular with many users, and they also offer useful information for creating a profile.
- User generated metadata, such as tags, comments, bookmarks, microblogs, playlists and collections that are often made in Web 2.0 platforms like Last.fm,

del.icio.us or Flickr.com can be aggregated and analysed to infer user's interests and preferences.

- User's bookmarks in a browser or RSS feeds in Google Reader, Netvibes or in any other RSS reader are another potential source of profile information. There is already a file format, OPML, that could be used gathering and transferring this data. The link destinations could then be analysed by utilising existing tags that can be found in some bookmarking applications like de.icio.us, or by the using some automatic methods to infer the topics of the links.
- Data from user's online calendars (e.g. Google Calendar) would be very useful for inferring interests and also future intentions. The type and topics of events are valuable information for interest analysis and this data also reveals locations which the user typically visits.
- User created content, such as blog entries could be analysed automatically to extract semantic metadata from them. Blogs and other user generated content is not necessarily easy for automatic analysis, because there may be a lot of creative usage of language, and sarcasm is used. Anyway, such automatic metadata suggestion could be a useful starting point for the user to manually correct and complement the profile.
- Social network: A social network, which is a graph of relationships within a group of individuals, are formed in many social media applications, such as weblogs and online forums. There is both explicit network information, such as lists of contacts or followers, and implicit information that can be used to construct the social network. Implicit information may be participation in same discussions, or using similar tags. Social relations between users indicate some similarities or common interests and provide potentially useful information for interest profiles as well. Social network can be used also for finding friends on a media site, and sharing the media experience with them.
- Usage: Queries and click-through-data from the interaction logs of a web service can be analysed and used to update a user's profile. Recently accessed resources will serve as implicit feedback to infer things that the user finds interesting at the moment.
- Media service providers could add such activity opportunities to their services that would both encourage users to create a user account and interact with content in such a way that produces information about the users' interests.
- If there is no existing data source to utilise, the final option is to let people create a user profile manually. Instead of free text descriptions, users could be offered a tool for creating profile with the help of semantic tags. Automatic tag suggestions based on external knowledge bases, such as Freebase, can be offered to help users to describe their interests. This approach allows freedom, but produces a semantic profile. Other tools with similar ideas could be created, for examples maps for giving location related information (e.g. home, work, summer cottage etc.).

The role and importance of personalisation depends on the context and aim of using media content. Users may be looking for some specific information or just generally browsing and looking for something new or interesting. The profiles are

most useful in the latter case when the user does not have a specific question in mind, or the user is pressed with time, and wants to see the most relevant news from her own perspective. Also, in the former case, the profile can also be used as the basis for suggesting additional items to view, and this way giving the user a reason to stay a bit longer at the media service.

Currently, news websites often let people see news prioritised from different perspectives, such as

- most recent as a combined list or grouped into traditional news categories (local, national, international, sports, business,…)
- most read within different time intervals
- most commented within different time intervals
- recommended by the editors

If a personal profile is available, it will be possible to offer content prioritised based on the user's personal interests.

In the discussions with the media companies participating in the CrossMedia project, the following use cases for user profiles have been identified:

- Hyper local news service
- Regional views into media content
- Event services personalisation
- Exploring news/media content guided by profile information.

These cases imply different information needs for the profile. In the hyper local personalisation, and particularly if the service is mobile, location, context and the most current intentions of the user are the key pieces of information. Hyper local and regional view can be regarded as being on the same continuum. In these both cases, person's location information plays the key role, but the profile can be utilised to prioritise what is presented to the user. The profile information can be used as an additional filter to raise or lower the relevance of individual pieces of news.

Event service personalisation is a good case for utilising personal profiles. Information generated in social media applications often relates to music, literature, films and hobbies, and events typically relate to these themes. Exploring news guided by profile information has potential particularly when there is a lot of information available. For example, each user's own interests as described in their profiles could be offered as a starting point for media content exploration. Also, in connection to searching large media archives, user profiles could be utilised rank to search results.

5. Current implementation

by Sari Vainikainen, VTT

This chapter describes the current implementation of the user profiling service and technologies used for creating a semantic profile. The service was implemented by utilising VTT's existing background as the starting point. The earlier components were updated to meet the requirement identified as important for Finnish media companies and circumstances. In the Pharos project the User Interest Profiler was an integrated part of the audiovisual search platform and the result of the profile analysis was stored into the platform database that was not hosted by VTT. In the Crossmedia project the profile service was developed for an independent service. This included development of the database. Other developed features include support for Finnish and Swedish tags, support for expressing explicit interests and exporting functionality of the profile. The main idea of the profile service is presented in Figure 1.

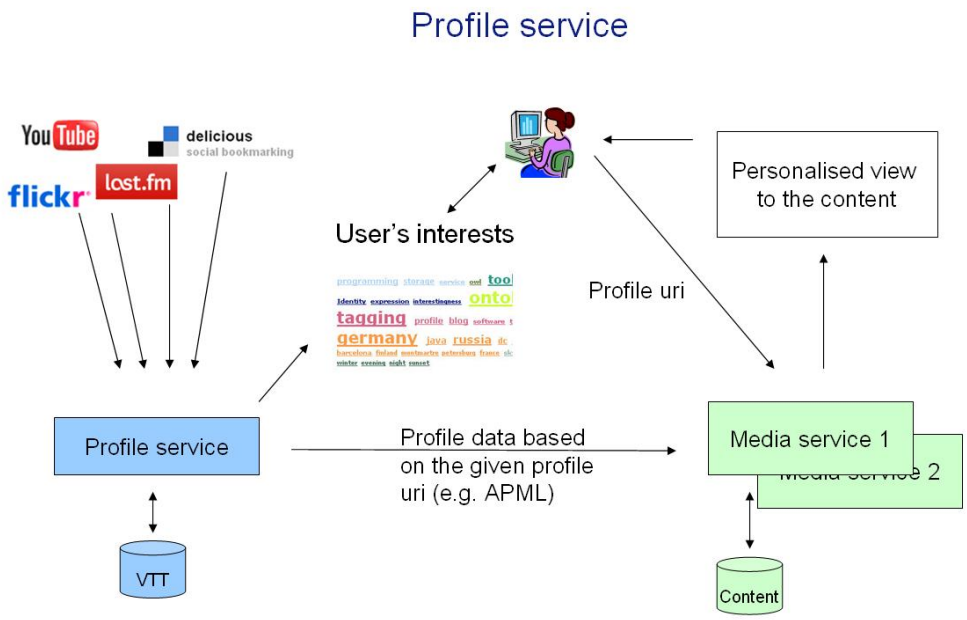

Figure 1. The profile service lets users create and maintain their profiles and export them to different media services. Media services can use this information for personalisation.

5.1. User interface functionality

The Profile service has as its core a component (background) that creates a tag based user's interest profile with the help of information available at user's different user accounts in existing social media services (del.icio.us, Flickr, YouTube, Last.fm) in the form of tags. The component uses the open API's of the

social media services to fetch the user's tags for the analysis. The semantic analysis (See chapter 5.2.1) also utilises information of tag usage and co-occurrence frequencies to clean user's tags and select the ones for a profile. The primary meaning for a tag is proposed and synonyms and categories that a tag belongs to are fetched and stored.

The profile is visualised for the user as a *tag cloud* and a *graph* (background). The size of the tag in the *tag cloud* implies the tag usage frequency and hence its importance to the user. When the user clicks a tag, a pop-over widget appears. The user can see and modify the inferred meaning, synonyms and related tags in the widget. The user can reduce or increase the relevance of a tag in the user's interest profile. Tags can be grouped in different ways such as usage frequency, source and categories. See Figure 2 for details.

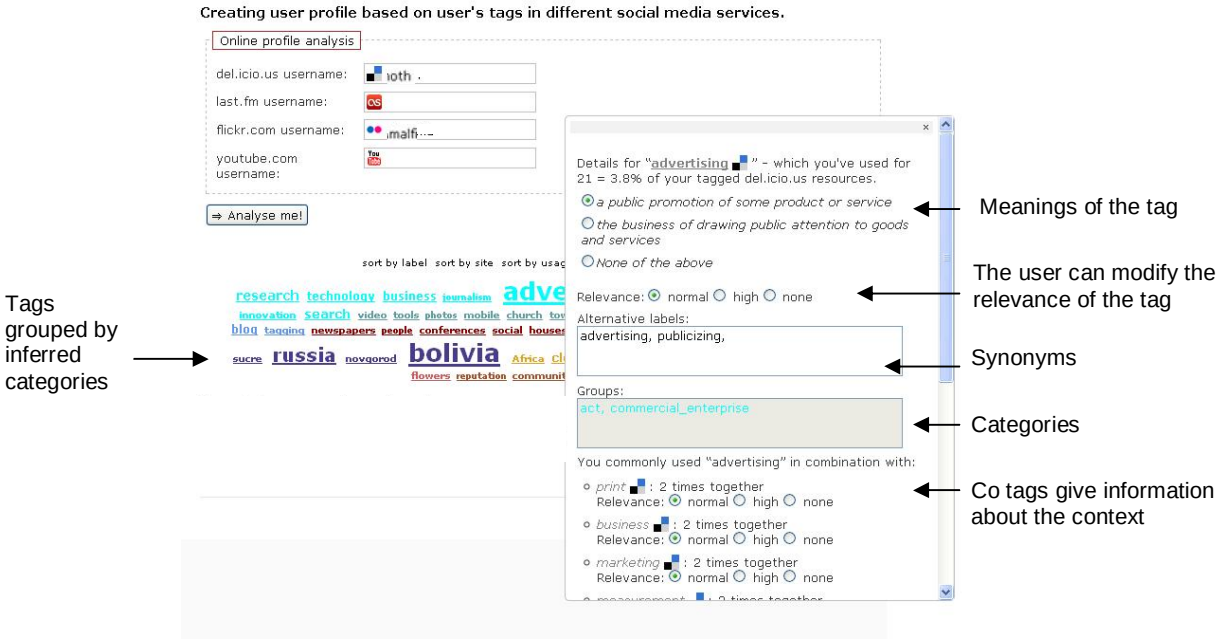

Figure 2 An example of the user interface of the profile service. The interest profile is visualised as a tag cloud and the user is able to modify the profile.

The *graph visualization* has the advantage that it visualizes the co-occurrences of tags. The node is the tag and edge between tags defines how many times the tags have been used together. By using the slide switch, the user is able to decrease or increase the number of tags included into her/his profile. The graph visualization is implemented as Java applet.

We also have *a semantic profile widget* (background), and with its help the user can define explicitly his/her interests such as favourite music, movies and hobbies. When the user starts to type an interest, suggestions coming from

ontology vocabularies are shown below the interest field. Different knowledge bases such as KOKO, Freebase and DBpedia are used for suggestions. These suggestions are restricted to certain topic, for example music includes suggestions from genre, artists and albums, in order to avoid overwhelming the user with choices.

An interest profile is built based on this information. The profile is available for exporting in APML format.

Users are able to use profile service without registration but when the user wants to save his/her interest profile, registration is required.

5.2. Technology

5.2.1. Semantic analysis of user's tags

The initial methods for semantic tag analysis using lexical analysis and semantic knowledge bases were developed by VTT in the Pharos project for the English language. We added the support for Finnish and Swedish tags in the CrossMedia project. The analyses use open APIs of the knowledge bases such as KOKO, DBpedia and Freebase. The reason to use several databases is that different knowledge bases contain different kinds of knowledge and not all information is available in one dataset. Users' tags can be very heterogeneous and this way we are more likely to find meanings for them.

As the first step, the analysis tries to recognize the language of the tag. Nouns are the most descriptive, and this way the most useful for an interest profile. Lexical analysis is used for cleaning the user's tags. Lexical analysis (word root, lexical class (noun, adjective…), singular/plurals) is made for the tag based on the recognised language. At the moment, the analysis supports lexical analysis for Finnish and English language tags. Swedish tags are handled in the analysis but they are not treated as smartly as in the other two languages, but this will be improved later, if the case study needs it. English and Finnish analysis includes also some misspelling corrections.

The analysis proceeds differently depending on the language. As the first step, Finnish and Swedish tags are mapped to KOKO concepts and after that, by using the English labels, to WordNet as well. DBpedia and Freebase are also used for mapping Finnish and Swedish tags because these knowledge bases support different language versions. An additional benefit of using DBpedia and Freebase is that they contain information relating to the concept, for example information relating to persons and locations.

English tags are analysed first with help of WordNet, which is both a dictionary and lexical database for the English language. By using linked data, relations to

other databases, such as DBpedia, are fetched. English tags are also mapped to KOKO.

As the result of the analysis we get alternative meanings for a tag. The analysis suggests a primary meaning for each tag, but each user is able to change the meanings of her tags, if the primary meaning was not correct. Information about the co-occurring tags is relevant for disambiguation purposes, because they often give hints about the context and meaning. Based on the selected meaning, additional information, such as synonyms, broader terms or other classification information and geo coordinates for location tags are fetched and stored to the database.

There are especially two areas in the analysis that need further development. One is to develop better methods for disambiguation. We have already earlier developed methods for disambiguation of English tags using the functionality of WordNet, but we need same kind of methods also for KOKO and DBpedia.

Also methods for creating reliable mappings between different knowledge bases need to be developed further. Linked data which provides owl:sameAs links between different databases is a valuable source and is utilised here, but it is not enough. Especially relating to KOKO we do not have this kind of mapping data available yet. As a result of the analysis we do produce skos:closeMatch relations for mappings but this is not directly reliable without checking.

This analysis is not limited only to tags but it can be used also in connection to semantic annotation of any textual content.

5.2.2. Semantic tagging widget

The Semantic tagging widget can be used to define user's interests, such as favourite music, video, hobbies, explicitly. When the user starts to type a tag, suggestions coming from ontology vocabularies are shown below the tag field. When user selects the suggested tag, its meaning is defined as an Uri to the database. Based on this, more knowledge can be accumulated. For example, if it is a location, its geo-coordinates, or if it is an artist, his music genre and other artists in the same genre can be fetched. The widget supports tag suggestions from different knowledge bases such as KOKO, Freebase, DBpedia and Geonames. Based on requirements of each case, the widget can be configured to use only one or several databases. It can also be configured to use new vocabularies and ontology, for example to support IPTC or vocabularies used in media companies (e.g. magazine vocabulary). It also supports restricting suggestions to certain topics.

The widget in itself is not restricted to creating only profile tags, but it can be used for annotating semantically any content.

5.2.3. Matching user profiles and content

At this stage of development we have focused on developing features for creating and analysing a profile. The profile is matched to different vocabularies including KOKO, Dbpedia, Freebase and WordNet, but better support for classification of interests needs to be developed. In earlier project we have used WordNet's super senses for classification, but this approach is not enough now that we have extended the analysis to better support other languages and knowledge bases. One relevant option is to use the YSA main classes and their KOKO mappings as a starting point.

For service providers, there are different opportunities to utilise the profile data in their services. They can use user's interests as such for querying their content based on free text indexing. These queries can be expanded with synonyms or other alternative labels provided by a profile.

The plan for the future is that the profile service will support different widely used vocabularies, such as IPTC. This means that if a service provider uses a supported vocabulary for describing the content, matching user profile data and content will be more straightforward. If a service provider uses its own vocabulary, mapping to ontology that is used to describe user profiles, needs to be done.

The first implementation of matching user profile data with the Magazine ontology and Event classifications will be made in the CrossMedia cases during Spring 2010. We can use the same semantic tag analysis methods that were developed for analysing user's tags for mapping service providers' vocabularies to profile data.

We have developed an initial version of the semantic annotation functionality for content in Finnish based on our semantic tag analysis methods. In cases where the service provider has very little structured metadata of their content, this functionality can be used to annotate content semi-automatically.

5.2.4. Exchanging profiles between services - Vocabularies and ontology

An important question relating to the concept of service independent profiles and an independent profiling service is, what kind of data needs to be transferred between the profile and media service providers and which vocabularies support describing this type of metadata. A user profile may consist of different parts such as user's interests, social network of friends or

contacts, demographic information (such as age, gender, family status), plans and intentions, preferences and behavioural patterns, which have different role and purpose for personalisation. At this point, we have focused on defining metadata relating to user's interests.

The data in user's profile includes the user's interests, relevance or importance of an interest to a user, the meaning of the interest and classifications of a user's interests. Other additional information such as different language versions, alternative labels for the same thing (synonyms, misspellings etc.), descriptions, relations of concepts, other knowledge that describes persons, places, music albums, films, companies are also relevant information for searching and matching content and user profiles.

Based on these requirements for the metadata, we have compared two approaches to define user's profile on structured format. One is APML XML format and other is combination of FOAF, SCOT and MOAT ontology.

APML

APML (Attention Profiling Markup Language) is an XML format that summarises user's interests (across multiple profiles) in a simple, portable way. APML is under development, but the goal is to define open standard for exchanging attention data between applications.

The main concepts are Implicit and Explicit interests and Source. APML v.1.0 (http://groups.google.com/group/apml-public/web/apml-1-0-draft-1) version extends features so that an Uri can be linked to a concept or that the user can manually provide more detailed information. Support for location coordinates has also been added. APML-RDF draft version is available at http://apml.pbworks.com/APML-RDF.

APML provides functionality to express users' explicit and implicit interests as tags, to create different profiles (work, home, etc.), to combine a user's interests from different sites and to define the relevance of interests as well as to express semantic meaning of a tag (v.1.0).

The developed profile service supports exporting the user profile in APML, but we extended APML to support expressing alternative labels, classification information, description and co-occurred tags (See Figure 3).

```
- <APML xmlns:vtt="urn:vtt-aplm-additions" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns="http://www.apml.org/apml-0.6" xsi:schemaLocation="http://www.apml.org/apml-0.6 file:apml.xsd" version="0.6">
  - <Head>
      <Title>Interests</Title>
      <Generator>VTTInterestExtractor v0.1</Generator>
      <UserEmail>foo@vtt.fi</UserEmail>
      <DateCreated>2009-11-12T10:21:23Z</DateCreated>
    </Head>
  - <Body defaultprofile="MainProfile">
    - <Profile name="MainProfile">
      - <ImplicitData>
        - <Concepts>
            <Concept vtt:groups="coding_system, communication" vtt:usedWith="article, example, java, test" vtt:description="(computer science) the symbolic
              arrangement of data or instructions in a computer program or the set of such instructions" vtt:altLabels="computer_code" xml:lang="en"
              rdf:about="http://www.w3.org/2006/03/wn/wn20/instances/synset-code-noun-3" updated="" from="VTTInterstAnalyzer" value="0.073171"
              key="code" />
            <Concept vtt:groups="communication, UNIX" vtt:usedWith="" vtt:description="an open-source version of the UNIX operating system" vtt:altLabels=""
              xml:lang="en" rdf:about="http://www.w3.org/2006/03/wn/wn20/instances/synset-Linux-noun-1" updated="" from="VTTInterstAnalyzer"
              value="0.073171" key="linux" />
            <Concept vtt:groups="European_country, location" vtt:usedWith="music, norway" vtt:description="republic in northern Europe; achieved
              independence from Russia in 1917" vtt:altLabels="Republic_of_Finland, Suomi" xml:lang="en"
              rdf:about="http://www.w3.org/2006/03/wn/wn20/instances/synset-Finland-noun-1" updated="" from="VTTInterstAnalyzer" value="0.048779998"
              key="finland" />
          </Concepts>
        - <Sources>
            <Source updated="2009-11-12T10:21:23Z" from="VTTInterstAnalyzer" type="application/rss+xml" name="del.icio.us public tag feed" value="1.00"
              key="del.icio.us" />
          </Sources>
        </ImplicitData>
        <ExplicitData />
      </Profile>
    </Body>
  </APML>
```


Figure 3 Example of the exportable profile as APML format.

FOAF, SCOT, MOAT

Another alternative format to describe a semantic user's tag based interest profile is to use FOAF ontology together with different tag ontology. Combination of FOAF, SCOT and MOAT ontology provides classes and properties to define versatile information related to an interest profile.

FOAF (The Friend of a Friend) ontology describes people, links between them and things they have created or done. FOAF ontology includes properties for describing different parts of the user profile such as basic information of a user (foaf:name, foaf:birthday, foaf:based_near, foaf:gender, foaf:workplaceHomepage etc.), user's interests (e.g. foaf:interest), social network (e.g. foaf:knows), a user's accounts (e.g. foaf:holdsAccount),projects and groups (e.g. foaf:member) and documents and images (e.g. foaf:made, foaf:topic).

When we need to extend profile data beyond interest profile, FOAF supports defining different types of user profile data. Many existing ontology (e.g. Tag ontology, SCOT and SIOC) are linked to FOAF, which provides additional expressivity. The most common way to reuse FOAF is to use foaf:Agent class for describing the entity that is responsible for an action.

SCOT (Social Semantic Cloud of Tags) ontology supports describing user's interests as scot:TagCloud, which contains information about the tags (scot:Tag) and their usage frequencies and co-occurrences (scot:Cooccurence) as well as of the total number of tags and posts (http://scot-project.org/ontology/) (See Figure 4).

SCOT supports creating of separate tag cloud of each site where the user has an account and also to define combined tag cloud (scot:composedOf). SCOT reuses FOAF (scot:createdBy), Tag ontology (tags:Tagging, tags:Tag, tags:Resource) and SIOC.

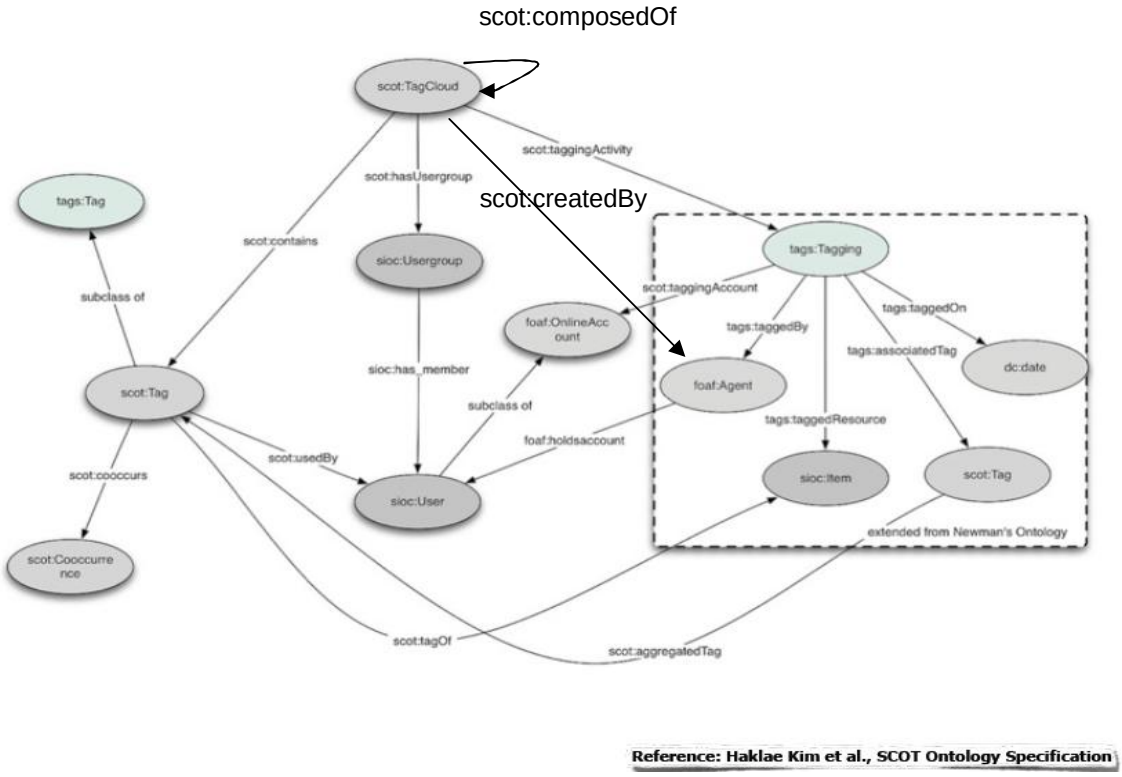

Figure 4 Core concepts and properties of SCOT ontology. Source: http://www.slideshare.net/echo4ngel/tag-ontologies

Because our goal is to create a semantic profile we need properties to define the meanings of interests. MOAT (Meaning of A Tag) ontology is developed for this purpose and it supports describing different meanings for one tag. The core concepts are Meaning and Tag (See Figure 5). MOAT reuses Tag ontology and FOAF.

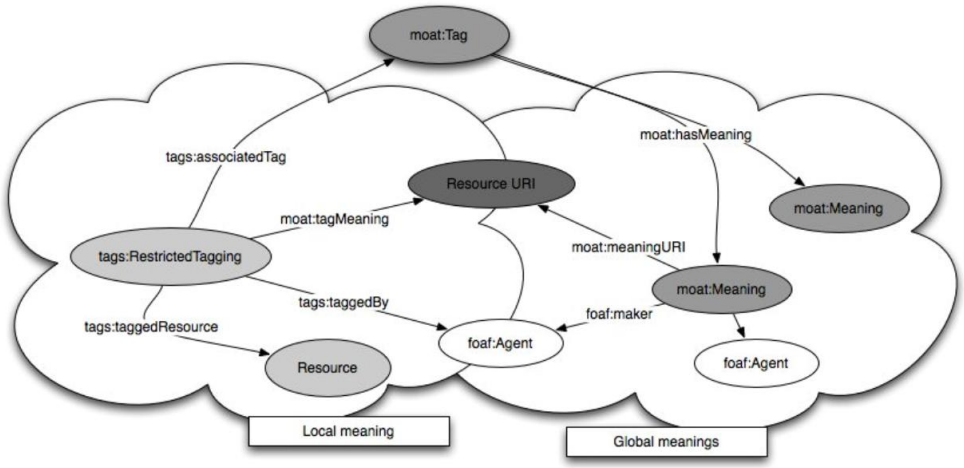

Figure 5 Core concepts and properties of Moat ontology. Source: http://www.slideshare.net/terraces/moat-from-tagging-to-semantic-web

In order to be able to utilise these ontology together, we needed to do some alignment of tag ontology. We defined moat:Tag as subclass of scot:Tag, which enables us to define the meanings of a tag more effectively.

As a summary these ontology provide functionality to:

- express users' interests as tags
- combine a user's interests as tag clouds from different sites and to define usage frequencies of tags
- connect this user's interest-related information with FOAF that can be utilised to express other aspects of a user profile as well
- express semantic meanings of a tag and relations between tags
- extend functionality with other ontology related to these (e.g. SIOC, SKOS)

Other vocabularies and ontology

Depending on the future requirements, service-independent profile service could support also other formats such as GUMO (the General User Model Ontology). GUMO has been created especially to support developing ubiquitous and user-adaptive systems. It is a wide ontology including user information from demographic data to personality, emotional and mental state. It also includes concepts and properties for defining context related information, user's interests, knowledge and preferences.

When the profile will be extended to other information than user's interests other vocabularies needs to be taken account. This includes for example vocabularies that define social network, context or contact information.

6. Conclusion

Digital cameras and mobile phones support automatic creation of geolocation information both by content creators and consumers, and this means that using geolocation metadata will be increasingly important in future editorial systems. Beside exact coordinates, geographical information includes other types of GIS primitives such as lineStrings and polygons that should also be supported in the future editorial systems. Geographical information can be used in editorial systems for determining the relevance of content to a user, for targeted advertisement and for expressing impact area of the story as well as for visualisation purposes.

User profiles are in key position to offer personalised content based on each user's interests. A big problem is how to make users share and maintain their profiles in different services. We have created the prototype of an independent profile service that lets users create and maintain their profiles for exporting them to different media services. Media services can use this information for personalisation. One basic idea in the current profile service prototype is that each user's activity in social media services can be used for creating a user profile. User activity in different services is also valuable potential source for keeping the profile updated.

We have developed methods for analysing user's interests based on the tags that they have created in the different services. Semantic analysis of user's tags was updated to support Finnish and Swedish tags in addition to English tags. Future development is needed for disambiguation of tags and for mapping concepts between different knowledge bases. We developed also some initial functionality for semantic annotation of content by utilising the same semantic analyses as for the tags. This can be used when developing methods for matching profiles and content. Exporting functionality of the profile as APML was added. Other options for exporting format were studied as well.

During spring 2010, methods for matching user profiles and content will be developed based on the requirements of the selected cases. The cases with KSF media and SanomaMagazine include user tests for getting feedback from the personalisation based on the profile service and media content. The cases will also give user feedback on the expected features of a service-independent profiling service.

Future editorial systems need to take into account supporting different opportunities for more personalised services. This includes handling user profile related data also in cases when the data is imported from external sources. It is also crucial that content metadata is on a level that supports various personalisation methods.