Personalised multimedia summaries

Cathy Dolbear[1], Paola Hobson[2], David Vallet[3], Miriam Fernández[3], Iván Cantador[3] and Pablo Castells[3]

[1] affiliation to be added
[2] Motorola Labs, UK. paola.hobson@motorola.com
[3] Universidad Autónoma de Madrid, Spain, {david.vallet,miriam.fernandez,ivan.cantador, pablo.castells}@uam.es

1 Introduction

In this chapter we will introduce an example of the application of semantic multimedia technologies to personalisation, and specifically to the creation of personalised summaries of multimedia content. Personalisation, in its simplest definition, is technology which enables a system to match available content, applications and user interaction modalities to a user's stated and learned preferences. In a multimedia context, the objective of personalisation is to enable content offerings to be closely targeted to the user's wishes, which can be achieved via methods such as content filtering, which selects content appropriate to a user's preferences from a set of available content, and recommendation which proposes content to a user based on various criteria which may include the user's previous acceptance of related content or on the consumption of related content by a peer group.

Personalisation is already well-known from content search domains e.g. the experimental aceMedia system in which personalised content is made available to an end user as part of a search operation (Evans, Fernandez, Vallet, and Castells 2006). This implies that the content contains some appropriate annotation, which may have been manually added or automatically generated. A personalised search application will transform user expressed (or learned) preferences into appropriate queries which can be matched to the available content metadata. Where content is not already annotated, analysis methods (such as those described in Chapter 6) can be used to derive metadata according appropriate schema or ontologies (as explained in Chapters 2 and 3).

The success or otherwise of personalisation depends partly on correctly understanding user expressed preferences and interpreting user actions to translate these latter into preferences, partly on sufficiently expressive and meaningful metadata being available in order to match the content to the user's requirements, and most importantly, on creation of appropriate reasoning technologies to enable the match to be made. The capture and representation of user preferences is an area known as user modelling, and is outside the scope of this book. The domain of metadata expression and representation is treated in chapters 2 and 3. The use of well defined and correctly specified annotation terms is very important, as it assists with interoperability of the reasoning tools which act on the content to achieve the desired personalisation. In other words, the personalised application examples which follow in this chapter rely on the availability of a unified and accurate set of metadata, which enable the appropriate content to be selected.

Personalised recommendations and personalised search applications have been extensively treated in the literature. Our focus in this chapter is processing of multimedia content to create personalised summaries which meet a user's need for a shortened version of the content, to meet requirements of time, interest level, and physical constraints such as storage space or available bandwidth. An example might be to produce five minutes' highlights from a premier league soccer match, for the busy soccer fan to view on a mobile device during their commute to work, or to share with friends when out socialising. Such a summary must not only meet time constraints (i.e. the user has specified how long the highlights can be), but must also have sufficient semantic meaning such that it can be understood as a self-contained unit. When we intend to personalise such summaries, we also must ensure that the selected content is that which the user is most interested in, and which generates an emotional response i.e. that the user feels that the chosen content has some specific meaning for them. However, many current multimedia content summarisation methods are signal based, and do not necessarily support semantic story telling according to a user's preferences and interests. Some examples are reviewed in section 2. We will also review user requirements for personalised summaries, and will describe some state of the art personalisation methods which can be applied to the personalised summarisation problem. In section 2 we will also introduce some high level usage scenarios, before moving on to a specific example of soccer matches summarisation in section 3. A complementary technique for contextualising personalisation, as a means to enhance the coherence of summaries, is presented in section 4.

2 Personalised multimedia summaries

Methods to automatically summarise long video sequences have been in development for some years, often with the objective of creating a short version of a long sequence to enable a user to review the material to determine if the full version would be relevant to them. Many authors have proposed the use of keyframes to represent a summary of a video sequence, for example (Chang, Sull and Lee 1999), in which the detected keyframes are intended to provide a compact representation of the video sequence. However, representation of a continuous video sequence by a series of still images does not convey the meaning of the video, but instead can only give a quick visual summary of some of the events in the sequence.

More recent work, such as (Sundaram and Chang 2001) and (Graves and Gong 2004), analyse the complexity and temporal action in the video content, in an attempt to generate more meaningful summaries which are composed of short temporal segments. In their experiments, the authors of (Sundaram and Chang 2001) summarised well-known movies down to as little as 5% of their original length (i.e. for a 165 minute movie, an 8 minute segment was created). Using a combination of automatic shot generation and manual selection, this method considers both visual complexity of the content, as determined by the amount of time it would take to understand the meaning of some content following a keyframe, and the film syntax i.e. how the producer arranges film shots to create the story. The approach taken by (Graves and Gong 2004) aims at creating an entirely automatic summarisation process, based on where the system determines that there is "action" in the video sequence. Where this would lead to useful results in applications where scene activity is the most important element e.g. surveillance video, it does not necessarily translate well to summarisation of other types of material such as TV drama or operatic performances, where people and objects may be stationary for long periods of time.

In the above methods, the objective is to produce a summary of the video content which enables a user to quickly understand the content without needing to view the entire sequence. This may be useful for applications where a user must review long sequences, such as, for example, many hours of surveillance video tapes or long TV programmes or movies from a video archive. The summarisation in these cases would aim to create a universally meaningful clip so that any user could understand the essence of the full sequence. This, however, is not applicable for applications where the summarised clip is intended to be the final item to be viewed by an end user, such as would be created for sports clips services. In such applications, the user subscribes to a service which will enable them to view summaries of their favourite events (e.g. a soccer or cricket match). These summaries are intended to be accessible via mobile communications networks, and should be available soon after the match is finished. Each user subscribed to the service has their own personal preferences for what they want the summary to contain, such as favourite team, player, etc, and it is unfeasible to create multiple personalised summaries using a manual method. Therefore we seek a method of automatic generation of sports highlights from a full length match according to the preferences of the fan, and in the next section, we will describe this method applied to soccer content.

3 Example application - personalised soccer summaries

The massive interest in soccer makes commercialisation of filtering techniques in this domain attractive and manually edited soccer highlights are already being marketed as a major application for third generation mobile phones. However, since substantial expertise and time is required to edit soccer highlights by hand, an advantage of an automatic system is to allow a user to receive personalised highlights. For instance, user requirements research (Evans 2003) has shown that fans wish to see specific events involving particular players, and are very keen on viewing summaries of soccer matches when mobile, as the key activities are only a small part of the game. A knowledge elicitation study with sports editors (Dolbear and Brady 2003) has shown that soccer highlights must tell the story of the match, and that flow of play and event causality, for example the events leading up to a goal, are very important elements of a good highlights package.

(Young 2000) notes that coherence comes from the selection of actions whose causal and temporal relationships highlight an underlying plot. User interaction, for example in automatic narrative generation tools, allowing the user to alter the state of the world at any given point in a story, can so radically alter the world that even the most accommodating plot lines cannot survive. This raises the question of how far we can personalise a summary before losing the sense of coherence. For example, including only events involving the soccer fan's favourite player may result in a meaningless sequence of disjoint events, providing no understanding to the viewer of what actually happened in the game. We need to make sure that personalisation only takes place within a framework of coherent summarisation, in order to avoid this problem.

This section describes the implementation of an automatic soccer highlights generation system (Dolbear 2004) and addresses the issue of how to measure users' satisfaction with the content they are provided with and how this relates to the length of summary they might pay for. We also investigate the trade-off between personalisation and coherence in a summary, by developing a novel quantitative measure of summary coherence based on the causal relationships between events.

3.1 Previous work

Soccer highlights, however, are summaries presented as the final product for user consumption. The information extraction problem has been addressed in the soccer domain using audio and video features such as colour density analysis, slow motion replay detection, penalty-box detection and speech-band energy to identify semantic events using machine learning techniques such as Bayesian Belief Networks (Ekin, Tekalp and Mehrotra 2003). With such systems, *any* event that can be recognised is deemed important enough to include in the summary. This leaves the generation of more meaningful summaries, containing only events relevant to a particular user, as an open area of research.

It has long been recognized in natural language processing research that an accurate summary includes all the narrative elements of the original text (Lehnert 1981), and the importance of a text unit depends directly on the number and quality of causal relations that the unit has to other text units (Trabasso and Sperry 1985). More recently, narrative coherence, modelled using tree-depth measurement in rhetorical structure theory trees, has been the basis for sentence selection algorithms for text summarisation (Mani, Bloedorn and Gates 1998). While the authors have shown that modelling textual coherence improves summarisation results, the causal relationships in all of these previous methods have been manually annotated. Our system is able to identify these causal relationships automatically, and we then use them to personalise the summaries.

From interviews with soccer fans asking them to rank events in priority order (Evans 2003), we have a clear idea of user preferences in the soccer domain. Goals were found to be the most important, followed by major referee decisions, sendings off, fouls, the build up to and celebrations following a goal, interviews with goal scorers or man of the match, and finally controversial incidents. This insight is used in the design of our user profile ontology in section 3.2.

A frequently used method for personalising a multimedia summary, for example (Ferman, Errico, Van Beek and Sezan 2002), is to assign a weight to each of the user's preferences, and use these weightings to vary the scores of the multimedia content entities, so that a resource allocation agent can then determine which content should be included in the personalised summary. An alternative is to use a collaborative filtering technique (Shardanand and Maes 1995). These are mainly employed in recommender systems providing personalised suggestions about items that a user may find interesting. Neighbourhoods of users with similar tastes (specified via user profiles) are formed and used to generate recommendations of items that a particular user may be interested in. Neither of these personalisation approaches allow for the *coherent* combination of a number of items into a summary, and so are insufficient for our purposes.

3.2 Method : Soccer ontology and 'neutral' summarisation

Since our primary focus is on information summarisation rather than extraction, we sidestep the need to extract information from the audio or video representation of soccer matches, and use the minute-by-minute "ticker-tape" reports widely available on many sports' websites.

While work has been done on information extraction from free-text soccer reports (Saggion, Cunningham, Maynard, Bontcheva, Hamza, Ursu and Wilks 2002) we avoid this complexity by using a template mining technique to extract information directly from text where there is an automatically recognizable pattern. For example on websites such as the BBC's, the number of event classes that are described is limited, so we can simply search for expected words and phrases such as "Goal", "by" (followed by a player's name) and "from left half". We use a soccer ontology containing 20 classes representing common soccer events such as Assist, Booking, Corner, Foul, and Goal. Each event class has a *start time*, *extra time, duration* and *player* property associated with it. Our training set consists of 126 examples of full length soccer match descriptions and their corresponding summaries, generated by manual annotation of those events shown in the highlights broadcast on television, which we use as a "ground truth" benchmark. Events are clustered into causally-related groups (which we term *context groups*). This is either done using the groupings chosen by the editor of the original web page (when the events are grouped into paragraphs on the web page) or they are clustered using a Markov chain to estimate the joint probability of those events occurring as a group. The transition probability matrix of the Markov chain is estimated by counting the frequency of occurrence of each event class and pairs of event classes. For example, P(Booking|Foul) is high, as many Fouls are followed by Bookings in the training set. To cluster events in the test problem into a group, events are added to the group in turn, and the joint probability of the group's occurrence is calculated. The group is terminated when this joint probability falls below a certain threshold. The relative priority of each context group is the probability of a context group being included in the summary, given that it has occurred in the full length sequence. This is estimated using a second Markov chain, whose transition probability matrix derives from the frequency of occurrence of events in the summaries of the training set (rather than full length sequences, as for the context group clustering). By introducing the concept of these context groups, we can generate a summary not consisting solely of disjoint, unrelated events, but which makes sense as a whole, and explains to the viewer, for example, what caused a player to be sent off the pitch, or how a goal came about.

Personalised summaries and user profile design

Figure 1 shows our user profile ontology, along with two instances, representing example users *Simon* and *Sarah*. The property values in these user profiles were chosen to reflect two users, one more interested in controversial events, and the other in skill and goal-related events. We bias the summaries for the two different users towards different narrative episodes by changing the priority of different context groups, based on their content. Since we know the duration of each event, we can limit the duration of the summary to the user's preferred length by choosing only the highest priority context groups until the required duration is reached.

User profile property	Instance 1	Instance 2
Name	Simon	Sarah
Summary length	60 seconds	5 minutes
Favourite club	Manchester City	Everton
Secondary clubs	-	Arsenal
Favourite player(s)	David Seaman, Nicolas Anelka	Wayne Rooney, Thierry Henry
Favourite event	Goal	Goal

Second favourite event	Sending off	Penalty
Third favourite event	Foul	Shot
Fourth favourite event	Penalty	Save
Fifth favourite event	Booking	Assist

Figure 1 Properties of the user profile class, along with values of two instances used in the personalisation experiment**s**

3.3 Results : Duration accuracy

The first question to answer is whether our system can produce summaries of the right length for different users. We varied the requested length of the summary between 30 seconds and 20 minutes. Then the duration error between this request and the actual summary output was measured for the 126 soccer matches, in a leave-one-out fashion. The experiment was first carried out using summaries based on the priorities of individual events, and then repeated using summaries where whole context groups of events were included at a time, so that the differences in the two methods could be evaluated. Figure 2 shows a graph of the mean percentage error in duration accuracy for different summary lengths, and it can be seen that single-event based summarisation is more accurate than context-group based summarisation, especially for shorter summaries since an event's duration is of finer granularity than a context group's. The mean duration of a single event is 19.2 seconds, compared with 27.1 seconds for a context-group. However, beyond about 300 seconds there is little advantage in using single-event based summaries, in terms of duration accuracy, and the advantage of context-group based summaries is in the additional coherence they provide to the overall summary.

Utility

To offer the user choices like, “We know you're an Everton supporter, would you pay for an extra five minutes to see Everton scoring from a penalty?” we introduce a measure of how well the content presented to the user fulfils their requirements, which we term *utility*.

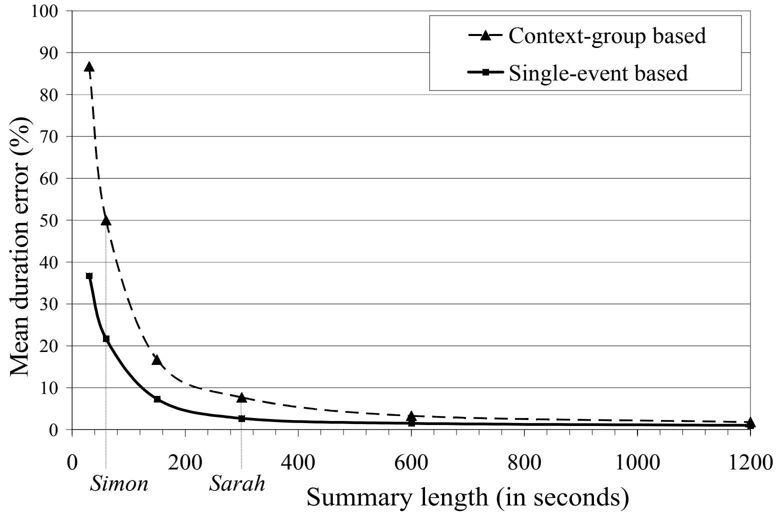

Figure 2 Personalised summary duration error against summary length, with the two users Simon and Sarah's preferred summary lengths marked.

Our utility function for a summary S and user profile U is defined as:

Error! Objects cannot be created from editing field codes.

where i is the index of user profile properties, and N is the number of properties in the user profile. The weightings w_i give higher priority to the more preferred events, and those involving a favourite player or club. Figure 3 shows how utility increases with summary length; for *Sarah* the rate of increase decreases with summary length, while for *Simon* it increases. *Simon* is a tougher customer to please than *Sarah*, although this difference is less noticeable at shorter summary lengths. This is because *Sarah*'s favourite events are included more often in the neutral (non-personalised) summaries than Simon's favourites.

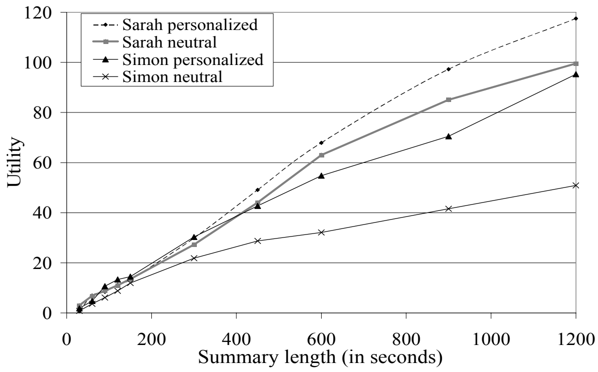

Figure 3 Personalised summary utility against summary length

Coherence

We now investigate the trade-off between coherence and personalisation. To what extent is our suggestion valid that constraining personalisation to the context group level improves coherence? Our coherence measure for a summary *S*, consisting of events E_t, E_{t-1},... E_1 is calculated as:

Error! Objects cannot be created from editing field codes.

That is, coherence is based on the causal relationships between the summary events, as calculated using the conditional probability of occurrence of the sequence, given the first event. Coherence is calculated here for summaries of the same length as the ground truth summaries, because coherence was found to decrease with summary length. Therefore, we do not vary the summary length for Simon and Sarah's preferences in this experiment in order to make useful comparisons of their coherence. Figure 4 shows the mean coherence of the summaries broadcast on television (the "ground truth"), compared with our neutral summaries generated using both single-event based and context-group based summarisation; and summaries personalised for *Simon* and *Sarah*.

Experiment	**Coherence**
Ground truth	0.112
Neutral, single event based	0.018
Simon, event-based, personalised	0.006
Sarah, event-based, personalised	0.009
Neutral, context-group based	0.117
Simon, context-group based, personalised	0.112
Sarah, context-group based, personalised	0.113

Figure 4 Mean coherence of various summaries; comparing ground-truth with neutral and personalised summaries

Figure 4 shows that coherence is much higher for context-group based summarisation than when single events are included in the summaries. The small difference in *Simon* and *Sarah'*s results can be attributed to the small variations in summary lengths. As suggested in (Young 2004), personalisation reduces coherence in the summary, but the drop is much smaller for the context-group based summaries than the single-event based ones, which shows the advantages of the context group idea in retaining coherence, even in a personalised summary.

3.4 Requirements for multimedia semantics in personalised summaries

In the context of a system for automatically generating personalised soccer highlights, we found that while single-event based summarisation has a smaller duration error than context-group based summarisation, since an event's duration is of finer granularity than a context group's, this advantage decreases significantly with summary length. The mean percentage error between the actual and preferred summary length also decreases as the summary length increases.

To entice a user to pay for extra content, or help them save time, we have developed a utility measure to quantify the additional benefit a particular user would gain from an increment in summary length. We found that utility increases with summary length, and that some users have higher utility than others, even for the neutral summaries. Finally, we investigated how our use of context groups of causally related events contributed to summary coherence. We found that including whole context groups in the summary, rather than just one event at a time, not only increased coherence, but mitigated the reduction in coherence due to personalisation.

Since the publication of Berners-Lee's vision of the Semantic Web in 2001 (Berners-Lee, Lassila and Hendler 2001), there has been increasing interest in the use of ontologies for describing the meaning of content, both on the World Wide Web and in databases. In the sense of the term "ontology" as understood by the semantic web community, the structures we have used to describe events and user profiles are little more than metadata. While it is understood that metadata should be standardised, and used as a syntactical exchange mechanism, the fundamental point of an ontology is that it is *not* standardised, but enables individuals to represent their own point of view via explicit semantics, forming part of the decentralised system of the Semantic Web. Extracting comprehensive semantics from multimedia, in order to populate an ontology expressed in Description Logics (using OWL-DL for example) is a long way off however. The semantic web community is moving towards understanding the need for provision of interfaces between standard W3C semantic web languages like RDF and OWL, and the type of processing that needs to be carried out for extraction of semantics from multimedia; or with other concrete domains such as spatial or temporal reasoning. It is at this juncture that a way forward may lie between the standardisation requirements of multimedia metadata and the more open, decentralised technology of semantic web ontologies.

4 Contextual coherence of personalised summaries

The previous sections address the achievement of intra-document coherence by analysing the relations between consecutive events in a personalised summary. However in a typical session, users view more than one story, and often several documents related to each one. E.g. for a sports day, one would want to watch important scenes of several matches, plus a summary of results, some analysis and discussion, interviews, etc. The amount of stories, daily events, and documents for each story, largely surpasses the available time of the idlest reader. Thus, it is not only important to summarise long documents to shorter versions, but to select a reasonable subset of relevant documents and topics to be presented to the user out of the massive flow of available news items, and compose the pieces (summarised or not) in an effective and coherent way, according to the user's particularities and live activity. In fact, in some cases the content is already delivered in a short format at the point of production (e.g. in news bulletins), and the summarisation need lies on the appropriate selection of items or segments. Compared with the intra-document perspective addressed in the previous section, the aggregation of summaries introduces a new dimension in the summarisation problem, where a larger variety of topics and a wider semantic heterogeneity are involved.

In this section we argue that it is possible to further enhance the coherence of composite summaries at the aggregative level, by analysing the relations between user preferences and the current, live user focus at runtime. Indeed, an important requirement in order for a personalised summary to be perceived as relevant and meaningful by the user is to improve its coherence with the ongoing course of user activities at the time the summary is generated. The idea of contextual personalisation, proposed here, addresses the fact that human preferences are multiple, heterogeneous, changing, even contradictory, and should be understood in context with the user goals and tasks at hand (Vallet, Castells, Fernández, Mylonas and Avrithis 2007).

Even if the user is believed to have a persistent set of user interests, either learnt by the system in the profiling phase, or manually provided by the user, it is assumed that such interests are not static, but vary with time and depend on the situation. In order to provide effective personalised summaries and develop intelligent personalisation algorithms, it is appropriate to not only consider a stable set of persistent user interests, but also to take into account the current user focus. Indeed, although users may have stable and recurrent overall preferences, not all of their interests are relevant all the time. Instead, usually only a subset is active at a given situation, and the rest can be considered as “noise” preferences. For instance, a user may enjoy documentaries about sea life, whereby the concept “sea” is important for her in the context of natural life documents, but this does not mean she is especially interested in sea battles when she is viewing a documentary about wars. Therefore, our model distinguishes a persistent component (which evolves at a slower pace) of a-priori user preferences, and a temporary, ad-hoc component, which is dependent on the live context within which the user engages in content retrieval tasks.

In our approach, the latter takes the form an explicit, dynamic representation of the live semantic context as a vector of weighted domain concepts, which is built by collecting ontology elements involved in user actions. This runtime representation of context is used in combination with the persistent user preferences in order to compute a focused, contextualised set of user interests. The computation of this set is achieved in two steps, consisting of a contextual expansion, followed by a contraction. In the first step, the initial preference and context sets are completed to form semantically coherent supersets, and in the contraction, a sort of intersection of the supersets is determined. This way, the semantic runtime context is used to activate different subsets of user interests at runtime, so as to achieve a coherence with the thematic scope of user actions, in such a way that out-of-context preferences are discarded. Finally, the contextualised user interests are used to achieve a better, more accurate and reliable personalisation of the retrieval results retrieved by the system in response to user queries.

Context is an increasingly common notion in Information Retrieval (IR) (Finkelstein, Gabrilovich, Matias, Rivlin, Solan, Wolfman, and Ruppin 2002). This is not surprising since it has been long acknowledged that the whole notion of relevance, at the core of IR, is strongly dependent on context – in fact it can hardly make sense out of it. However, context is a difficult notion to grasp and capture in a software system. In our approach, we focus our efforts on this major topic for content search and retrieval systems, by restricting it to the notion of semantic runtime context. The latter can be defined as the background themes under which user activities occur within a given unit of time. In this view, the problems to be addressed include how to represent the context, how to determine it at runtime, and how to use it to influence the activation of user preferences, contextualise them and predict or take into account the drift of preferences over time (short and long term). In our current solution to these problems, the runtime context is represented as (is approximated by) a set of weighted concepts from a domain ontology. For example, if a user is querying and reading about ecologic damages in a certain region, the context may be made of domain concepts such as *fire*, *toxic spills*, *air*, *river*, *fauna*, etc.

Our approach to the contextual activation of preferences is then based on a computation of the semantic distance between each concept in persistent user preferences and the set of concepts in the current context. This distance is assessed in terms of the number and length of the semantic paths linking preferences to context, across the semantic network defined by the ontology. This can be expressed as:

$$\mathrm{CP}(u,y) = \mathrm{P}(u,y) \cdot \bigcup_{x \in \mathcal{O}, x \xrightarrow{p} y \text{ in } \mathcal{O}} \mathrm{C}(x) \cdot w(p) \qquad (1)$$

where the union symbol denotes the algebraic sum (i.e. $a \cup b = a + b - a \cdot b$), $\mathrm{P}(u,y) \in [0,1]$ is the intensity of interest by a user u for a concept y of a domain ontology $\mathcal{O}$, $\mathrm{C}(x) \in [0,1]$ is the degree of importance of a concept x in the current context, $x \xrightarrow{p} y$ denotes there is a path p of semantic relations $r_i(x_i,x_{i+1})$, $i = 1, \ldots, k$-1, connecting x to y in the ontology (i.e. $x_1 = x$, $x_k = y$), and $w(p)$ is the propagation power of each path, which depends on the semantic strength (a value in [0,1]) assigned to the semantic relations that make up the path, namely $w(p) = \prod_{i=1}^{k-1} w(r_i(x_i, x_{i+1}))$.

For instance, in the previous example, a steady user interest for sea life would be found to be in context with the retrieval session, provided that the domain ontology includes e.g. *oil spill* as a special case of ecologic accident with impact on sea life, whereas a user preference for e.g. some basketball team would be out of place given the current focus of user activity. If, say, *fishing industry* was also in the context, the relevance of the user interest for *sea life* would be intensified, if the fact that the fishing industry *depends on* sea life is a known semantic relation in the ontology (see Figure 5). For further details on this method, the reader is referred to (Vallet et al 2007).

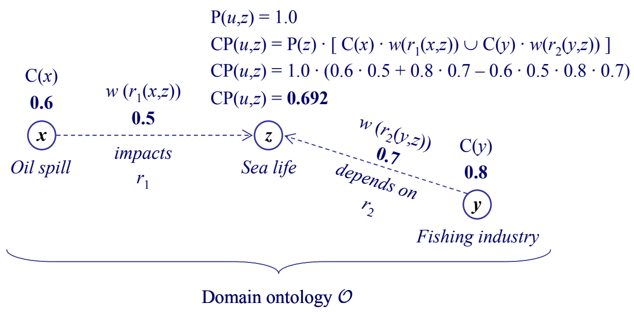

Figure 5 Contextual value of a user interest for "*sea life*" with respect to "*oil spill*" and "*fishing industry*" in a retrieval session

Ultimately, the perceived effect of contextualisation is that user interests that are out of focus for a given context are disregarded, and only those that are in the semantic scope of the ongoing user activity (a sort of intersection between user preferences and runtime context) are considered for personalisation. This would mean that, for instance, information about the damage of an oil spill on the sea life would have higher priority than the economic impacts, when a personalised summary about the accident is built. In practice, the inclusion or exclusion of preferences is not binary, but instead ranges on a continuum scale, where the contextual weight of a preference decreases monotonically with the semantic distance between the preference and the context, as determined by equation (1).

The extraction and inclusion of real-time contextual information as a means to enhance the effectiveness and reliability of long-term personalisation enables a more realistic approximation to the highly dynamic and contextual nature of user preferences. The ontology-driven representation of the domain of discourse, proposed in the previous sections, provides enriched descriptions of the semantics involved in retrieval actions and preferences, and enabling the definition of effective means to relate user preferences and context. The gain in accuracy and expressiveness obtained from an ontology-based approach brings additional improvements in terms of retrieval performance.

The proposed contextualisation technique brings a clearer benefit in multi-document (or multi-topic) summarisation, involving the automatic selection of a subset of available multimedia documents (or document segments on a variety of subjects). For instance a personalised multi-document summary may include a list of clips of several sports events, biased towards the user's favourite sports, teams, players, etc. This could include a couple of (summaries of) soccer matches, a basketball match, and a golf tournament. If the user pays more attention to the golf clips, the effect of contextualisation would consist of the summary automatically reorganising itself by increasing the space devoted to golf contents. This would be a consequence of temporarily raising the a-priori (persistent, long-term) user preference for golf, taking into account the ongoing user actions (semantic runtime context). This temporary, focused profile is what we are calling a contextualised user profile. The advantage of contextualisation are obviously higher when the initial multi-document spans across a wider subject range (e.g. including politics, sports, culture, etc.).

The contextualisation technique has been implemented in an experimental prototype, and tested on a medium-scale corpus. The latter consists of 145,316 multimedia documents (445MB) from the CNN web site (http://dmoz.org/News/Online Archives/CNN.com), annotated with the KIM domain ontology and KB (Kiryakov, Popov, Terziev, Manov, and Ognyanoff 2004), publicly available as part of the KIM Platform, developed by Ontotext Lab, with minor extensions. For the experiment, we have built a testbed including ten hypothetic context situations (scenarios), each consisting of a sequence of user actions defined a priori, including queries and clicks on summary items, detailed step by step.

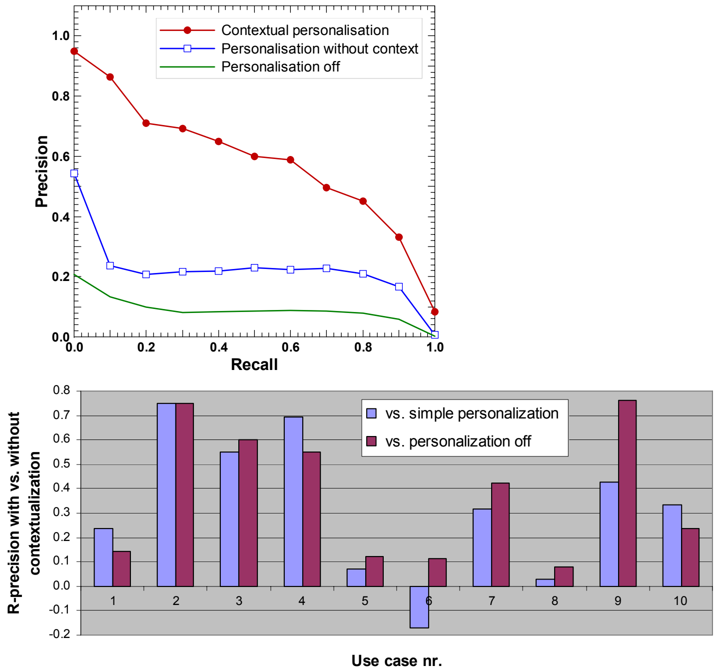

Figure 6 Comparative performance of personalised search with and without contextualisation, showing the average precision vs. recall curve (left), and the comparative precision histogram (right), for ten scenarios

The results of this experiment are shown in Figure 6, comparing the performance of contextual presonalisation vs. personalisation alone, and no personalisation. It can be observed that the contextualisation technique consistently results in better performance with respect to simple personalisation. The experiment shows how the contextualisation approach significantly enhances personalisation by removing out-of-context user interests, and leaving the ones that are indeed relevant in the ongoing course of action.

5 Conclusions

To be added

References

Berners-Lee T., Lassila O. and Hendler J. (2001) The Semantic Web. Scientific American May 2001

Chang, H.S., Sull, S. and Lee, S.U. (1999) Efficient video indexing scheme for content-based retrieval, IEEE Transactions on Circuits and Systems for Video Technology, **9** (8), pp 1269-1279

Dolbear C. and Brady J. M. (2003) Soccer Highlights Generation using A Priori Semantic Knowledge *IEE International Conference on Visual Information Engineering*, VIE 2003, University of Surrey, UK. July 2003.

Dolbear C. (2004) Personalised Information Filtering using Event Causality. DPhil thesis, University of Oxford 2004

Ekin A., Tekalp A.M. and Mehrotra R. (2003) Automatic Soccer Video Analysis and Summarization, *IEEE Transactions on Image Processing*, **12**(7), pp 796--807.

Evans A. (2003) User-Centred Design of a Mobile Football Video Database, *Second International Conference on Mobile and Ubiquitous Multimedia*, Norrkoping, Sweden, Dec 2003

Evans, A., Fernandez, M., Vallet, D. and Castells, P. (2006) Adaptive multimedia access: from user needs to semantic personalisation, IEEE International Symposium on Circuits and Systems, ISCAS 2006, Kos, Greece, 21-24 May 2006

Ferman A.M., Errico J. H., Van Beek P. and Sezan M. I (2002) Content-based filtering and personalisation using structured metadata *Proceedings of the Joint Conference on Digital Libraries* p393, ACM Portland, Oregon. July 2002.

Finkelstein L., Gabrilovich E., Matias Y., Rivlin E., Solan Z., Wolfman G. and Ruppin E. (2002) Placing Search in Context: The Concept Revisited. ACM Transaction on Information Systems **20** (1), 2002, pp. 116-131.

Graves A. and Gong S. (2004) Wavelet-based holistic sequence descriptor for generating video summaries, British Machine Vision Conference (BMVC 2004), Kingston, UK, 7 - 9 Sept 2004

Kiryakov A., Popov B., Terziev I., Manov D., and Ognyanoff D. (2004) Semantic Annotation, Indexing, and Retrieval, Journal of Web Semantics **2** (1), pp. 49-79.

Lehnert W.G. (1981) Plot Units: A Narrative Summarization Strategy *Cognitive Science* vol 4, pp 293-331 1981

Mani I., Bloedorn E. and Gates B. (1998) Using cohesion and coherence models for text summarisation *Proceedings of the AAAI Spring Symposium on Intelligence Text Summarisation*, pp 69-76 1998.

Saggion H., Cunningham H., Maynard D., Bontcheva K., Hamza O., Ursu C. and Wilks Y. (2002) Extracting Information for Information Indexing of Multimedia Material *Third Language Resources and Evaluation Conference* 2002

Shardanand U. and Maes P. (1995) Social Information Filtering: Algorithms for Automating "Word of Mouth", Proceedings of Computer Human Interaction CHI'95 p 210-217 1995

Sundaram H. and Chang S. F. (2001) Condensing Computable Scenes using Visual Complexity and Film Syntax Analysis. *Proceedings of the International Conference on Multimedia and Expo, ICME'01*. IEEE Tokyo, Japan, August 2001.

Trabasso T. and Sperry L. L. (1985) Causal relatedness and importance of story events *Journal of Memory and Langugage* **24** (5) pp 595-611

Vallet D., Castells P., Fernández M., Mylonas P. and Avrithis Y. (2007) Personalized Content Retrieval in Context Using Ontological Knowledge. IEEE Transactions on Circuits and Systems for Video Technology, special issue on the convergence of knowledge engineering, semantics and signal processing in audiovisual information retrieval (in press)

Young R. M. (2000) Creating Interactive Narrative Structures: The Potential for AI Approaches *AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment* AAAI, Stanford, CA, March 2000.