

Query Independent Measures of Annotation and Annotator Impact

James Lanagan and Alan F. Smeaton
Clarity: Centre for Sensor Web Technologies
Dublin City University
Glasnevin, Dublin 9, IRELAND
{jlanagan,asmeaton}@computing.dcu.ie

ABSTRACT

The modern-day web-user plays a far more active role in the creation of content for the web as a whole. In this paper we present Annoby, a free-text annotation system built to give users a more interactive experience of the events of the Rugby World Cup 2007. Annotations can be used for query-independent ranking of both the annotations and the original recorded video footage (or documents) which has been annotated, based on the social interactions of a community of users. We present two algorithms, AuthorRank and MessageRank, designed to take advantage of these interactions so as to provide a means of ranking documents by their social impact.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Information Filtering, Selection Process

General Terms

Algorithms, Human Factors, Measurement

Keywords

Annotation, Information Filtering, Data Quality

1. INTRODUCTION

The modern-day web-user plays an increasingly active role in the creation of content for the web as a whole. The large-scale adoption by internet sites of the technologies collectively referred to as web 2.0 has enabled users to create and upload more and more content to the web directly. As a result, information on a webpage or site may come from many unregulated sources with no specific information about its integrity, trustworthiness or authority. We provide a means of qualifying these sources, creating a ranking of the authors of new content based not only on the authors themselves, but also their interactions with the user community as a whole.

Based on the corpora of annotations created during previous experiments [5] as well as the system discussed in Section 3, we have developed two algorithms designed to incorporate the annotations of authors, along with the authority of the authors themselves, into the judgement of importance and significance of user-created content within corpora of original documents. The aim is to show that through the incorporation of ‘annotations’, it is possible to improve the ranking of papers which should be considered relevant to users’ information needs. These algorithms are based on the theoretical framework of data quality; we consider this as applied to the federated data of user-generated content.

This paper is laid out as follows; in Section 2 we discuss the motivations for the creation of an in-context social annotation system; our Annoby annotation system is described in Section 3; Section 4 gives the theoretical grounding for the algorithms that we have developed; finally, we reveal the algorithms that we have created which are designed to take advantage of user generated annotations in Section 5.

2. RELATED WORK

Adding an opinion to the bottom of a blog post or news story is something which internet users have come to take for granted, especially since the creation of resources the likes of Youtube ¹ and Digg ² which allow for more than just commenting; in the case of Digg, user comments are collected so as to promote the commented-upon item up to prominence on the front page of the Digg website. This philosophy of boosting based on social-impact is one which we share; with so much content being generated, it is necessary to form new ways in which to rank and filter content so as to provide a benefit and not hindrance to the user.

RefferalWeb [3] was a first attempt to bring together retrieval and social networks as the social aspects of information retrieval had not been properly leveraged before then. Collaborative search and filtering systems have been utilising similar techniques to recommend items of interest to people based on like-minded individuals’ past choices [4, 8].

The usefulness of comments and annotations themselves has also been studied in the past. Annotation forms a bridge between the separate activities of reading and writing, allowing the reader to take a more active role in the creation and dissemination of information. In the physical world annotations take the form of underlining, marginalia, highlighting etc. with the exact method specific to each annotator. The

¹<http://youtube.com>

²<http://www.digg.com>

vast majority of these annotations are anchored to specific points (phrases, words, paragraphs) within the source documents, due mainly to the increased effort required on the part of the annotator to recreate the context of disassociated annotations [2].

Tools enabling ubiquitous digital annotation of the web are available [9] providing free annotation of web-pages, and removing the restrictions of commenting only where specific functionality has been provided. Instead a culture of “active reading” [1] is fostered which helps to blur the line between writing and reading.

3. ANNOBY

Annoby brings together 3 currently separate aspects of sports recording. Users are able to read match reports, view the match video associated with the reports and create in-context anchored comments which are used as the basis for discussion between users of the system. The immediate and easy access to both visual and written media, coupled with the ability to leave comments within the text for other users, leads to a more directed and communal style of annotation. The novelty of the system is the opportunities it provides to become up-to-date with any talking points and also to contribute easily to on-going discussion.

The Rugby World Cup 2007 was chosen for its huge appeal; sport has always been a highly contentious topic of conversation with each person having his/her own opinions about the events that take place. Annoby is designed to capture this conversation and present it to the community of users.

3.1 Architecture

Figure 1 illustrates the production process which the sports media goes through before being presented to the user, a process similar to that of [5]. Information comes from two main sources before being gathered into a single match record: video recorded from television and web reports taken from newspaper web-sites. The initial video recording was made in MPEG-1 and later transcoded to MPEG-4, a step necessary to allow for streaming video playback through the Darwin Streaming Server³. Post-processing of the recorded video is done so as to remove all non-game footage such as studio discussions. The written reports are taken from two newspapers chosen for the differing view-points and perspectives they present; the Guardian Unlimited⁴ and the Irish Times⁵.

We used the *Cut_detect* algorithm [6] for shot cut detection on the MPEG-1 video files. The initial shot boundaries formed the skeleton onto which event segments were fitted. Once shot boundaries had been detected, we calculated the per-second confidence values for event boundaries of the entire video. Event detection was based on the work of [7], however a full description of the manner in which events were detected is beyond the scope of this paper. Per-second confidence values are used to calculate the highest-valued event segments within the video. A 20 second event-window performed well at combining highly rated shots which belong to the same event segment, taking the middle keyframe of this event to represent the event. All information about event

³<http://dss.macosforge.org/>

⁴<http://sport.guardian.co.uk/>

⁵<http://www.irishtimes.com/sports/>

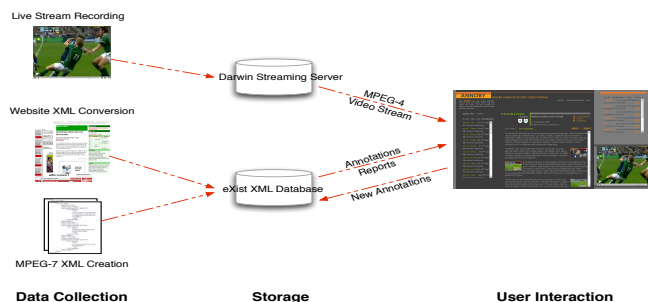


Figure 1: Annoby System Architecture

times and keyframes associated with the video are stored in MPEG-7. All data files required by the system are in fact stored in XML format within an eXist XML database⁶ which provides the required functionality for the storage and querying of XML documents.

The use of a dynamic threshold was an important step in retrieving the most ‘exciting’ events from a match video. We chose to restrict the number of events returned for a video to the top 10. Some games did not have 10 highly exciting events and so the dynamic threshold was combined with a lower bound so as not to return meaningless segments.

Upon logging in to Annoby, the most recently uploaded game is presented by default with the assumption that this game will be the current point-of-focus for users, and therefore where the most annotation will have taken place since a user last logged in (Figure 2). On the left of the screen the list of games is also available for easy navigation between matches; to the right of the screen, a list of all the commentators on the current game. The main focus of the interface is the centre of the screen, where keyframes representing the most exciting events are presented in-context within the reports. Video playback is to the right of the reports. The details of the match are shown above the main reports in the centre of the screen, along with comment statistics providing an indication of how much conversation has been taking place about the game.

Keyframes of the most significant events are presented inline within the report so as to make the connection between the media types more explicit. The position of keyframes alternated between left and right, each keyframe being presented at approximately the same offset into the report as corresponded to the percentage time into the match. Sports reports are theoretically written with the first paragraph summarising the entire game, and events are then presented in chronological order providing an outline of the match as a whole. Using this fact and without any semantic analysis, we can present video events in the region of their corresponding text description.

The purpose of Annoby is to create a corpus of annotations/comments on which to test the algorithms discussed in Section 5. While comments are of great use, allowing users to feel a part of the on-going discussion, they should not be intrusive and make the reading experience worse. Large numbers of comments may prove disorientating and disruptive to the flow of a report. While the option to show all comments is present via the “Show All” button, single

⁶<http://exist.sourceforge.net/>



Figure 2: The Annoby main interface

comments may be shown by clicking on the small orange annotation symbol, an example of which may be seen before the second paragraph in Figure 2. It is also possible to highlight the comments of a particular commentator (Figure 3), allowing users to more easily find the new comments or comments which most interest them.



Figure 3: Highlighted comments by a particular user

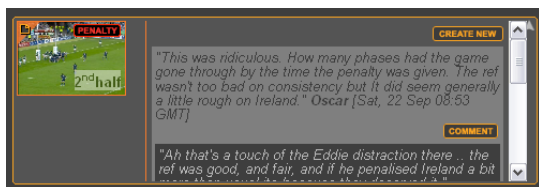


Figure 4: Annotation of event keyframes

With the combination of keyframes and reports, the ability of annotating a keyframe and implicitly annotating the event which it represents was introduced (Figure 4). An un-annotated keyframe is present at the bottom of Figure 2 showing a grey border. A grey border is used along with a grey annotation symbol in the top left corner to signify the absence of an annotation, but the opportunity to create annotations remains. In order to provide consistency, annotations on both event keyframes and free-text are presented in orange. The use of tags on each important event within the video allows users to see at a glance what event the keyframe represents. Clicking on the image directly however will result in the playback of the video for the event.

4. DATA QUALITY AND USER GENERATED CONTENT

Within the context of Annoby we have attempted to look at the quality of the data being provided by a federated web data source of users. The approach of [11] fits nicely with the ideas we have for providing an automatic quality measure to the contributions/annotations of users of the web. Zhu & Gauch give 6 attributes on which the quality of web-pages (and implicitly the data within those web-pages) may be judged. We shall adapt their idea of quality measures for an entire web-page to take into account instead the annotations provided to a web-page. We do not consider the *relevance* attribute:

- **Currency:** "How recently a web-page has been updated, measured as the time stamp of the last modification of the document." — how long has an annotation been in the system? What is the timespan of an entire thread?
- **Availability:** "Calculated as the number of broken links on a web-page divided by the total numbers of links it contains." — not really considered in this work but may be relevant if private annotations were to be considered.
- **Information-to-Noise:** "The proportion of useful information contained in a Web page of a given size meaning the ratio of the total length of the tokens after preprocessing divided by the original size of the document." — this same definition holds.
- **Authority:** "The reputation of the organization that produced the Web page based on the Yahoo! Internet Life reviews⁷." — who has created this annotation and what is their influence and standing within the community of users? This is provided by AuthorRank.
- **Popularity:** "How many other web-pages point to this particular web-page? — how many replies has this annotation received, and who has been replying?"

5. ALGORITHMS

We have developed two techniques which we use to improve the ranking of documents provided in response to a user's information need. The re-ranking is query independent and may be used as a guide to help a user's browsing. As our basis for quality we take the theoretical basis provided by [11], with the exception of relevance and availability. The main premise of the following two equations is that importance flows from commentator to annotation to document. If annotations are made by authors who are important or influential within the network of authors, then the value of the article should be increased. We may also say that the value of a comment or annotation is dependent on its author.

Each author receives a score based on the annotations which they have created. AuthorRank then allows us to decide which authors should be considered most expert or most likely to have promoted the supplementary creation of information useful to the user community as a whole. By focussing on the conversations between the top ranked authors, we can find documents which are both most likely to satisfy the user's needs, and which also are most likely to serve as the anchor for informative and insightful annotations.

AuthorRank: AuthorRank takes into account three different characteristics of an author's interactions with the network; the amount the author writes; the level of interaction that the author has with the rest of the community; and

⁷<http://www.zdnet.com/yil>

the level of influence which the author has over the conversation being had. These factors are combined within equation (1). The first of these, Avg_{wc} , is the average amount (a word count) that the author has written per annotation.

$$A_R = \frac{1}{3} * [Avg_{wc} + \frac{1}{1+\beta} * \{ \frac{S_T + \alpha * S_B}{S_{TOT}} + \beta * [\frac{R_T + \gamma * R_B}{R_{TOT}}] \}] + [\frac{ts}{Avg_r}] * [\sum_{x=1}^n \frac{d_x}{2^x}] \quad (1)$$

The central part of AuthorRank takes into account the *cohesiveness* of the author by looking at the percentage of annotations which are the start/head of a thread, S , versus those which are replies to other annotations, R . Annotations are further divided into those annotations which have received replies, and those which remain barren (have had no replies). The down-weightings α and γ are applied to annotations which have remained barren, β is the down-weighting applied to replies.

The depth of annotations (the position of the annotations within a thread of annotation) made by an author, d_x , is taken into account giving an idea of the author's catalytic potential to create conversation. One other aspect of AuthorRank worth noting is its attempt to discern how argumentative or provocative the author is. This is done through the judgement of activity within a thread containing the author's annotation. A thread with low average response time, Avg_r , and a high life-span, ts , may be judge as a "hot topic".

MessageRank: While the AuthorRank of equation (1) reflects the global characteristics of each author, equation (2) gives the MessageRank of each particular annotation. This rank is affected by the AuthorRank, A_R , of the author who created it, the replies it receives, the depth at which it is found within a thread, and the AuthorRank of authors involved in the thread in which the annotation is found.

$$M_R = \frac{1}{3} * [A_R + \frac{\tau}{1 + \tau} * \{ \frac{M_w^2}{T_w * T_{Med_r}} * [T_i - M_d] \}] + [\sum_{x=1}^n \frac{A_{R_x}}{d_x}] \quad (2)$$

The size of annotations in terms of words, M_w , gives the first indication of its impact. Longer messages are considered more important as there is a greater probability of these messages will stimulate further conversation. We also take into account the number of words, T_w , within the entire thread in which the annotation is found. In order to judge the influence of the annotation on its containing thread, the average word count of annotations within the thread, T_{Med_r} , must be calculated. By taking into account the depth of the thread, T_i , as well as the depth at which the annotation is found, M_d , increased importance is given to annotations which are found higher (or earlier) in longer threads. Annotations from a thread which contains many entries are considered to be more interesting or important by virtue of the fact that more people are interested in the conversation being had [10].

In some contexts, news or discussion forums say, a long thread between just two authors may be thought of as a type of "flame war" where the value of the information provided by the authors involved is likely to degrade as the dialogue continues. We therefore take into account the number of authors found within the thread, as well as who exactly these authors are. To account for topic drift or change of focus, the influence or strength of the interactions between the author of an annotation and the author, A_{R_x} , of any other anno-

tation in the current thread is proportional to the distance between the two authors, d_x , within the thread.

6. FUTURE WORK

Annotations help a user to focus on the most pertinent points being made; through the application of the AuthorRank and MessageRank algorithms, we hope to show that the inclusion of annotations does indeed provide a better solution to addressing a user's information needs.

Acknowledgments

This work was funded by Science Foundation Ireland as part of the CLARITY CSET, under grant number 07/CE/I1147.

7. REFERENCES

- [1] M. Adler. *How to Read a Book*. Touchstone, 1972.
- [2] A. Brush, D. Barger, J. Grudin, A. Borning, and A. Gupta. Supporting Interaction Outside of Class: Anchored Discussions vs. Discussion Boards. *Proceedings of ACM CHI2002*, 2002.
- [3] H. Kautz, B. Selman, and M. Shah. Referral web: Combining social networks and collaborative filtering. *Communications of the ACM*, 40(3):63–65, 1997.
- [4] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, 40(3):77–87, 1997.
- [5] J. Lanagan and A. Smeaton. Sportsanno: What do you think? *RIAO'2007: Proceedings of the 8th conference on Information Retrieval and its Applications*, 2007.
- [6] C. O'Toole, A. F. Smeaton, N. Murphy, and S. Marlow. Evaluation of Automatic Shot Boundary Detection on a Large Video Test Suite. In *CIR'99 - The Challenge of Image Retrieval: 2nd UK Conference on Image Retrieval*, 1999.
- [7] D. Sadlier and N. O'Connor. Event Detection in Field Sports Video Using Audio-Visual Features and a Support Vector Machine. *IEEE Transaction on Circuits and Systems for Video Technology*, 15(10):1225, 2005.
- [8] B. Smyth, J. Freyne, M. Coyle, P. Briggs, and E. Balfe. I-SPY-Anonymous, Community-Based Personalization by Collaborative Meta-Search. *Proceedings of the 23rd SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 367–380, 2003.
- [9] M. Wilson, D. Daniels, and J. Phillips. Annozilla, 2004.
- [10] W. Xi, J. Lind, and E. Brill. Learning Effective Ranking Functions For Newsgroup Search. *Proceedings of the 27th annual international conference on Research and development in information retrieval*, pages 394–401, 2004.
- [11] X. Zhu and S. Gauch. Incorporating Quality Metrics in Centralized/Distributed Information Retrieval On The World Wide Web. In *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 288–295, New York, NY, USA, 2000. ACM.