

Towards Truly Ubiquitous Life Annotation

Ashley Smith, Wendy Hall, Hugh Glaser and Leslie Carr

School of Electronics and Computer Science

University of Southampton

{ads04r, wh, hg, lac}@ecs.soton.ac.uk

Abstract

Throughout the age of information technology, we have experienced our lives getting simpler through the use of computers. Computer technology is becoming more and more integrated into our everyday routines, and as the technology becomes available we come to rely upon it. In this document we present a discussion of current technology that may be used to annotate the life of a human being, and propose a future in which diaries are obsolete in favour of computer technology and the Semantic Web.

1 Introduction

Annotation is the process of appending descriptive meta-data to existing information. This can be as simple as adding comments to a program listing, or as intricate as storing geographical co-ordinates in a JPEG image file. Crucially, an annotation describes a resource without actually duplicating or modifying it. For this reason, it is common for indexing services and search engines to generate and store annotations of content. While it is possible to run a search of a PC's hard disk for all files matching specific criteria, if the files are annotated in a particular way, in many cases the search engine can query the annotations to narrow down the search before searching a much smaller subset of the data. The same system is employed in the real world. If we were searching for a restaurant in a large city, we would not wander aimlessly around the city until we came to an establishment that fits our requirements. Instead, we look in a directory full of descriptions that are sorted in a logical way, and the description points us to the place for which we are looking.

In this respect, memory is an annotation of a life. As we live out our daily routines, we do not store the things that happen to us in our heads, we merely store references to past events. These may be recalled at a later date, but not actually re-lived. The brain stores that which it judges important, and does not worry about the small details. Using technology, it may not only be possible to duplicate this behaviour, but also fill in the gaps that are initially left out.

2 The Myth of Too Much Information

The Semantic Web [1] is a web of knowledge that surrounds the existing World Wide Web. The Web, though a rich source of information, is of no use to a machine as it is designed with human readers in mind. The Semantic Web contains annotations in the Resource Description Framework (RDF). This framework is a very simple, well-defined collection of facts in the form of 'triples'.

In its present form, the Semantic Web is capable of describing web resources such as documents and images, but also such things as people and projects. For example, when annotating a photograph in the Semantic Web, there are two vocabularies that prove very useful; the WGS84 vocabulary¹ for describing geographical positions (latitude/longitude pairs) and the Dublin Core vocabulary² which allows, among other things, time and date descriptions. Using the Friend of a Friend³ vocabulary it is also possible to determine that a photograph depicts another resource, such as a person.

But a photograph also depicts an event in a life. The Semantic Web allows for logical inference, so we can assume if a person took a photo, and we know where the person was when the photo was taken, we can assume that the photograph was taken at that point, even if the photograph contains no geographical meta-data itself. Reversing this, if we know a person took a photo at a particular location at a certain point in time, then the person was actually at that position at the time the photo was taken [4]. This suggests to us that provided we store as much meta-data as we possibly can, even if we do not have all the information we want, we can deduce the rest from other sources. But it is the information that lies there dormant waiting to be discovered that often goes un-noticed.

3 Pervasive Data Collection

Whether intentionally or unintentionally, we leave electronic 'footprints' wherever we go [2]. Ideally, we would like some way of keeping and annotating these for future use. In the modern world, where affordable data storage

¹http://www.w3.org/2003/01/geo/wgs84_pos#

²<http://purl.org/dc/elements/1.1/>

³<http://www.foaf-project.org/>

capacity is increasing there is no reason not to collect as much data as possible. Websites such as Last.fm⁴ exist to log the listening habits of music fans, and digital television systems have the ability to keep track of what their users watch, so that they may recommend similar material based on the user's preferences. The cost of keeping these logs is small considering the potential benefits, and the data may yet have some future purpose that we do not yet know. This is a common fault of human memory; we remember that which is important at the time, which is not always what is important later.

Yet, any attempt to automate life annotation should be completely automatic. The findings of a 2001 study by Rodden et al [3] indicate that although categorisation, sorting and indexing (all forms of annotation in various ways) are desirable by most people, very few are actually willing to spend any effort in order to achieve this ideal.

We therefore propose a personal data store that indiscriminately collects as much information as possible about the life of a human, with no concern as to how the information is to be used.

4 The Current System

The current implementation takes the form of a data store that runs on a desktop computer, which employs various 'data collector' applets that generate or collect meta-data, as well as an application that runs on a handheld device such as a PDA or a Smart Phone. These portable devices have the advantage that people generally use them anyway, and relying on them for data collection does not place any requirements on the user to change his/her way of life. Currently the handheld application is capable of logging the user's contacts and calendar appointments, the device's status properties, and the user's current geographical position if the device is fitted with a GPS receiver. The application runs on a Windows Mobile-based PDA.

The desktop application takes the PDA logs and inserts the data into a local RDF knowledge store. Additionally, the application is always active, running in the background. It can be set to run any number of tasks at certain times of the day or at regular intervals. These tasks are known as 'data collectors', and are pre-compiled executable programs which perform a specified task and return information in an RDF format. The main application, after running one of these collectors, imports the returned data into the knowledge store alongside the data collected from the handheld. Collectors may be added or removed at any time, and may be written in any programming language. Typically these collectors will return information that is available via the web or some local process, such as local weather or traffic reports, but can be anything that is of use. We may also want to collect more general information, that is relevant not just to the known user but universally recognised knowledge, such as the geographical co-ordinates of major cities or landmarks. After all, a user is more

likely to ask "when was the last time I visited the University of Southampton" than "when was the last time I visited 50.9372, -1.39787". If the store contains this knowledge, we can use it to our advantage, combining it with our personal data to gain a much richer description of the user's life.

5 Conclusions and Future Work

The real value of this system has yet to be discovered. Our aim is to log as much information as possible without ever deleting it, in case it becomes useful in the future. The current implementation annotates the life of a user, and therefore provides an aid to memory using nothing but off-the-shelf hardware. Additionally, other than installing the software on a PC and/or PDA, it requires the user to perform no extra tasks that interfere with lifestyle whatsoever, which is important if widespread adoption is to be encouraged.

This system works because an increasing amount of data is becoming available to a computer. People store their appointments in digital organisers rather than paper diaries, and much correspondence is by email, so it is not hard to scan a subject's inbox to determine their contacts and how often they interact. It is even possible to store a user's movie and music taste digitally and automatically, due to the increasing dependence on general computer hardware for playing multimedia.

Currently, a test subject is using the system and has amassed almost a year's worth of life annotations. The next step in our work is to evaluate this data in an attempt to determine exactly how valuable it is. If use can be obtained from data collected using commercially available hardware, we believe that as hardware becomes more advanced, we can collect more and more information about a life, possibly to the point where diaries are all but obsolete and biographies can be written by a computer at the touch of a button.

References

- [1] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, pages 35–43, May 2001.
- [2] Hugh Glaser. Semantic squirrels. In *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*, Athens, Georgia, USA, November 2006.
- [3] Kerry Rodden, Wojciech Basalaj, David Sinclair, and Kenneth Wood. Does organisation by similarity assist image browsing? In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 190–197, Seattle, Washington, April 2001.
- [4] Ashley Smith, Leslie Carr, and Wendy Hall. An opportunistic approach to adding value to a photograph collection. In *Proceedings of the Fourth International Semantic Web Conference*, Galway, Ireland, November 2005.

⁴<http://www.last.fm/>