Sensor Nets Discover Search

Keynote Talk

Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies
School of Computing
Dublin City University, Ireland

Alan.Smeaton@dcu.ie

ABSTRACT

In the world of information discovery there are several major trends which are emerging. These include the fact that the nature of search itself is changing because our information needs are themselves becoming more complex and the data volume is increasing. Other trends are that information is increasingly being aggregated, and that search is now becoming information discovery. In this presentation I address a different kind of information source to the usual media, scientific, leisure, and entertainment information we usually consume, whose availability is now upon us, namely data gathered from sensors. This covers both the physical sensors around us which monitor our environment, our wellbeing and our activities, as well as the online sensors which monitor and track things happening elsewhere in the work and to which we have access. These sensor information sources are noisy, errorsome, unpredictable and dynamic, exactly like both our real and our virtual worlds. Several wide-ranging *sensor web* applications are used to demonstrate the importance of event processing in managing information discovery from the sensor web.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; I.2.9 [**Artificial Intelligence**]: RoboticsÑSensors; H.4.2 [**Information Systems Applications**]: Types of SystemsÑDecision support

General Terms

Algorithms, Experimentation, Human Factors, Measurement

Keywords

Sensor networks, information seeking, event detection

Summary

We hear a lot these days about how search, and information seeking in general, is changing. Partly this is due to the increasing volume and complexity of information, the ease

with which we can lay our hands on it, the increasing complexity of tools for managing information as well as our own increasingly complex information needs. The basic search or retrieval paradigm on which the most popular search services is based, has been fine for almost all that we needed in the past and indeed is still adequate for most of what we need today, but changes to this paradigm for information access are necessary to keep pace with these demands.

We also hear a lot about how our search results are moving up in terms of sophistication with news and current affairs being a good example. Instead of visiting news sites to gather together individual news stories which may be of interest to us but mostly will not, we use news aggregators like Google News to bundle and group things together, aggregating stories into events which are the units retrieved for us. Search results can also be clustered by similarity or some other criterion so that we can quickly grasp an overview of our search output.

Lastly, we also hear much about how information seeking is changing to be information *discovery* rather than search. Continuing the example of how news spreads, we're increasingly finding that a news event will live, and will die, based on its own value or merit to its readers, and we can now pick up our news from sources like Twitter and Facebook where the re-tweets will propagate, or kill off, a news item. Here we have a working case where information really does find us rather than we find the information, replacing the model of a user having an information need explicated as a query, using an information retrieval system to match the query against the *documents* and retrieving those most likely to be of relevance. Of course a huge amount of effort has gone into making this an effective matching operation — modeling the retrieval process mathematically, accommodating relevance feedback by introducing state-based systems, modeling the user's context, modeling links and dependencies between documents, and so on.

In the case of managing multimedia information, because the *documents* are not discrete objects but are compositional, there are even further complications and then of course the task of processing multimedia and normalising its content into an accurate representation is itself a huge challenge.

So what we can observe from current trends is that search is changing, information is increasingly aggregated, and search is becoming discovery. What does the future look like?

In this presentation I address a different kind of information source whose availability is now upon us, namely

data gathered from sensors. This covers the physical sensors around us (environment, place, motion, activities like traffic, weather, people movement, crowd gatherings like concerts and sports) as well as the online sensors we have access to (blogs, tweets, news etc.). Often termed the *sensor web*, this information source is characterised as being noisy, errorsome, unpredictable and dynamic, exactly like both the real and the virtual worlds in which we live, work and play. I begin by introducing several wide-ranging sensor web applications to show the breadth and pervasive nature of the sensor web across our lives and I then address how

information in the sensor web is being, and will be, managed. From this it will be clear that events ...detecting, characterising, representing, comparing, searching, ranking and ultimately pushing events, is a key element and again I use examples to illustrate.

Acknowledgments