working stack called Rime. The Rime stack implements sensor network protocols ranging from reliable data collection and best-effort network flooding to multi-hop bulk data transfer and data dissemination. IP packets are tunnelled over multi-hop routing via the Rime stack.

Interaction with a network of Contiki sensors can be achieved with a Web browser, a text-based shell interface, or dedicated software that stores and displays collected sensor data. The text-based shell interface is inspired by the Unix command shell but provides special commands for sensor network interaction and sensing.

To provide a long sensor network lifetime, it is crucial to control and reduce the power consumption of each sensor node. Contiki provides a software-based power profiling mechanism that keeps track of the energy expenditure of each sensor node. Being software-based, the mechanism allows power profiling at the network scale without any additional hardware. Contiki's power profiling mechanism is used both as a research tool for experimental evaluation of sensor network protocols, and as a way to estimate the lifetime of a network of sensors.

Contiki provides a flash-based file system, called Coffee, for storing data inside the sensor network. The file system allows multiple files to coexist on the same physical on-board flash memory and has a performance that is close to the raw data throughput of the flash chip.

To ease software development and debugging, Contiki provides three simulation environments: the MSPsim emulator, the Cooja cross-layer network simulator, and the Netsim process-level simulator. The development process for software for Contiki typically goes through all three simulation stages before the software runs on the target hardware.

In October 2008, major industry players Cisco and Atmel joined Contiki. Cisco, Atmel and SICS jointly announced uIPv6, the world's smallest fully compliant IPv6 stack. uIPv6 builds on the uIP stack and is integrated in Contiki.

The Contiki team currently consists of sixteen developers from SICS, SAP AG, Cisco, Atmel, NewAE and TU Munich.

Adam Dunkels from SICS, Sweden, is the winner of the 2008 Cor Baayen Award for a promising young researcher in computer science and applied mathematics.
http://www.ercim.org/activity/cor-baayen

Link:
http://www.sics.se/contiki/

Please contact:
Adam Dunkels
SICS, Sweden
Tel: +46 70 773 1614
E-mail: adam@sics.se

Preferential Text Classification: Learning Algorithms and Evaluation Measures

by Fabio Aiolli, Riccardo Cardin, Fabrizio Sebastiani and Alessandro Sperduti

Researchers from ISTI-CNR, Pisa and from the Department of Pure and Applied Mathematics at the University of Padova, are explicitly attacking the document classification problem of distinguishing primary from secondary classes by using 'preferential learning' technology.

In many contexts in which textual documents are labelled with thematic classes, a distinction is made between the primary and secondary classes to which a given document belongs. The primary classes of a document represent the topic(s) that are central to the document, or that the document is mainly about. The secondary classes instead represent topics that are somehow touched upon, albeit peripherally, and do not represent the main thrust of the document.

This distinction has been neglected in text classification (TC) research. We contend that it is important and deserves to be explicitly tackled since, in most contexts in which the distinction is made, the degree of importance of a misclassification can depend on whether it involves a primary or a secondary class. For instance, when a patent application is submitted to the European Patent Office (EPO), a primary class from the International Patent Classification (IPC) scheme is attached to the application, and that class determines the expert examiner who will be in charge of evaluating the application. Secondary classes are attached only for the purpose of identifying related prior art, since the appointed examiner will need to determine the novelty of the proposed invention against existing patents classified under either the primary or any of the secondary classes. Thus, for the purposes of the EPO, failing to recognize the true primary class of a document is a more serious mistake than failing to recognize a true secondary class. Similar considerations apply to other scenarios in which the distinction is made.

In a concerted attempt to address this distinction, we define preferential text classification, a task which we define as the attribution to a textual document d of a partial ordering among the set of classes C. This partial ordering specifies whether or not a given class 'applies more than' (or 'is preferred to') another class in the document. In particular, we focus on a special case of preferential TC; namely, the case in which each document is associated to a 'three-layered' partial order. This consists of a top layer of one or more primary classes, each of which is preferred to those in a middle layer of secondary classes, which are in turn each preferred to those in a bottom layer of 'non-classes' (ie classes that do not apply at all to the document).

The original contribution of our work is twofold. First, we propose an evalua-

tion measure for preferential TC, in which different kinds of misclassifications involving either primary or secondary classes have a different impact on effectiveness. Second, we attack preferential TC by using a learning model, dubbed the Generalized Preference Learning Model, that was explicitly devised for learning from training data expressed in preferential form, ie in the form "class c' is preferred to class c" for document d". This model allows us to draw a fine distinction between primary and secondary classes in both the testing and learning phases, thus making use of the different importance of primary and secondary classes to which a training document belongs. Experiments run on WIPO-alpha, a well-known benchmark dataset consisting of manually classified patents, show that the Generalized Preference Learning Model outperforms standard (ie non-preferential) state-of-the-art learning approaches.

Link:
http://www.isti.cnr.it/People/F.Sebastiani/Publications/IRJ08b.pdf

Please contact:
Fabrizio Sebastiani
ISTI-CNR, Italy
Tel: +39 050 3152 892
E-mail: fabrizio.sebastiani@isti.cnr.it

DataCell: Exploiting the Power of Relational Databases for Efficient Stream Processing

by Erietta Liarou and Martin Kersten

Designed for complex event processing, DataCell is a research prototype database system in the area of sensor stream systems. Under development at CWI, it belongs to the MonetDB database system family. CWI researchers innovatively built a stream engine directly on top of a database kernel, thus exploiting and merging technologies from the stream world and the rich area of database literature. The results are very promising.

Rather than simply transmitting the raw measured data, current state-of-the-art sensors are capable of a limited amount of processing. This feature has many positive effects, such as keeping the network usage and costs as low as possible. However, this is not enough to replace the role of well-equipped nodes that gather streaming sensor data from multiple sources and which account for the biggest share of the processing cost. These nodes should be able to perform complex query processing on large amounts of incoming data, meeting strict real-time deadlines even in periods when the frequency of incoming data explodes.

Our work focuses on this part of the sensor research. We are designing and developing a system called the DataCell, which is capable of efficiently collecting and processing high volumes of stream data. We are currently studying the DataCell over the stream application scenario of an ambient home setting. The DataCell is positioned as a data refinery cell that acts as an easily programmable data hub in a multi-network environment. Its task is to collect, filter and aggregate information from different sources to enable complex decision making, such as control of the lighting based on audio/video presentations. The challenge in an ambient environment is to hide the computer from the casual user, even while it is actively steering the environment. An example query in the ambient scenario could be, "tune the

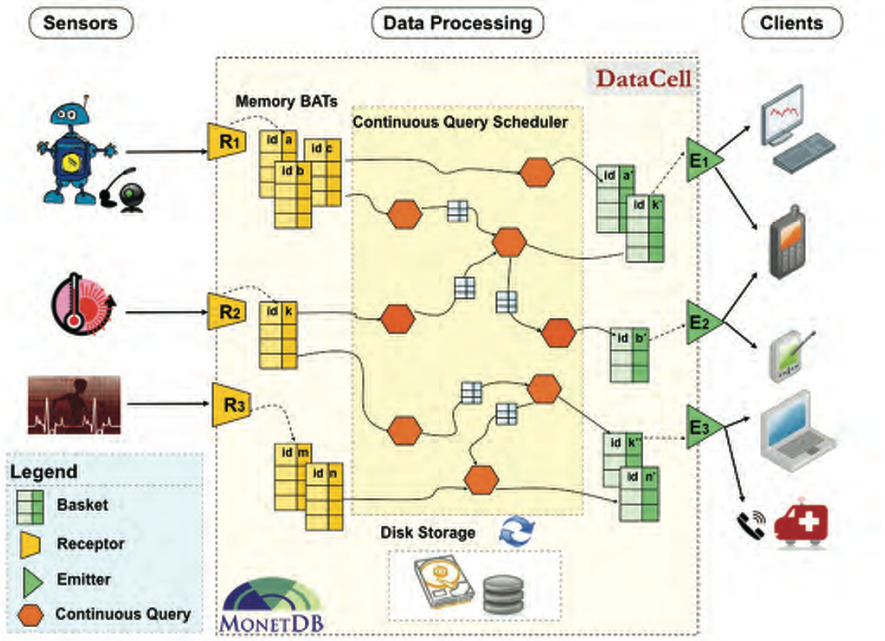

Figure 1: The DataCell in the ambient scenario.