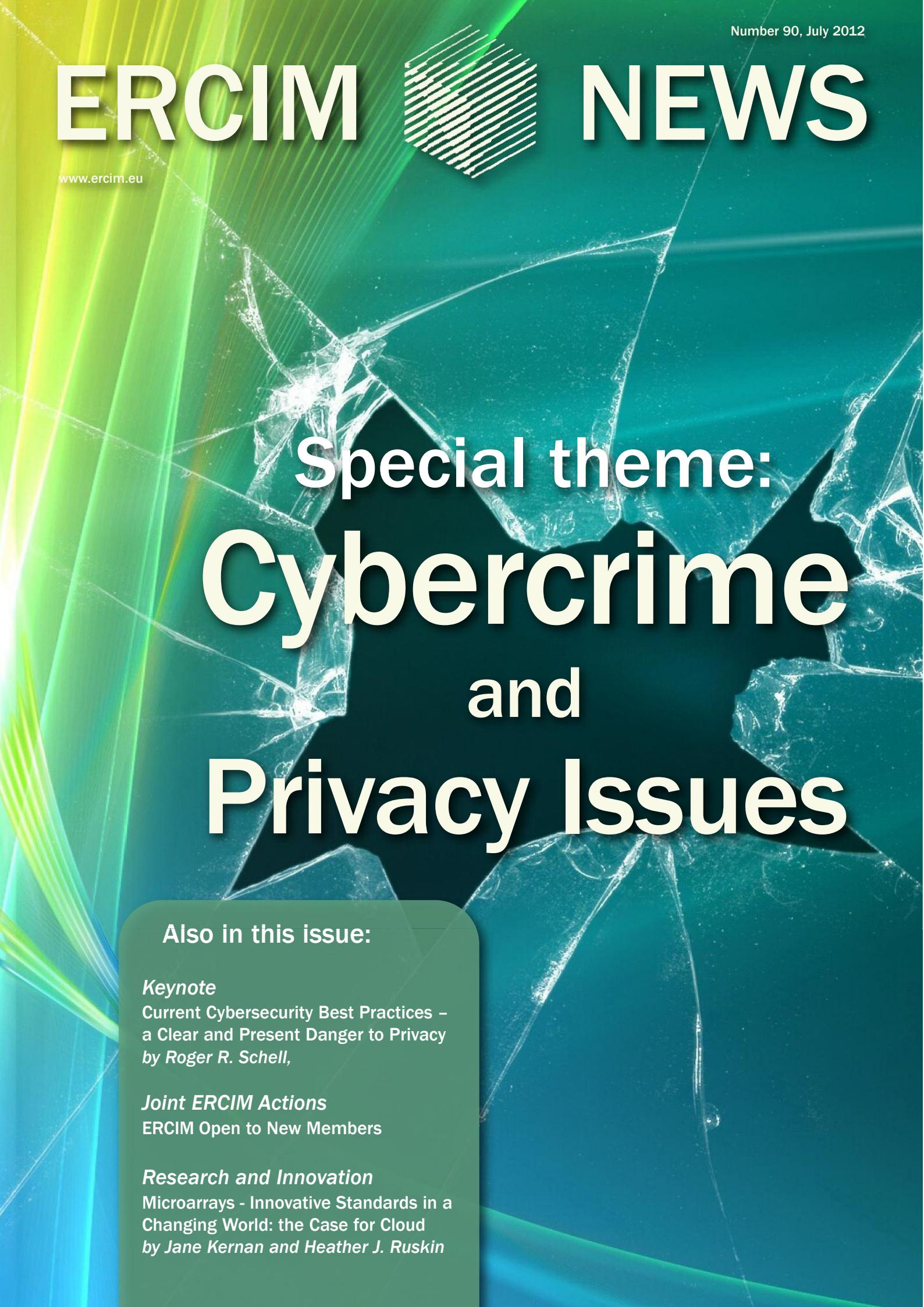
Number 90, July 2012

ERCIM NEWS

www.ercim.eu



Special theme:

Cybercrime and Privacy Issues

Also in this issue:

ERCIM News is the magazine of ERCIM. Published quarterly, it reports on joint actions of the ERCIM partners, and aims to reflect the contribution made by ERCIM to the European Community in Information Technology and Applied Mathematics. Through short articles and news items, it provides a forum for the exchange of information between the institutes and also with the wider scientific community. This issue has a circulation of about 8,500 copies. The printed version of ERCIM News has a production cost of €8 per copy. Subscription is currently available free of charge.

ERCIM News is published by ERCIM EEIG
BP 93, F-06902 Sophia Antipolis Cedex, France
Tel: +33 4 9238 5010, E-mail: contact@ercim.eu
Director: Jérôme Chailloux
ISSN 0926-4981

Contributions
Contributions must be submitted to the local editor of your country

Advertising
For current advertising rates and conditions, see
http://ercim-news.ercim.eu/ or contact peter.kunz@ercim.eu

ERCIM News online edition
The online edition is published at http://ercim-news.ercim.eu/

Subscription
Subscribe to ERCIM News by sending email to
en-subscriptions@ercim.eu or by filling out the form at the
ERCIM News website: http://ercim-news.ercim.eu/

Next issue
October 2012, Special theme:
What is computation? Alan Turing's Legacy

Roger R. Schell, President of ÆSec, founding Deputy Director of the (now) US National Computer Security Center. He is considered as the "father" of the Trusted Computer System Evaluation Criteria (the famous "Orange Book")

Current Cybersecurity Best Practices – a Clear and Present Danger to Privacy

Not only is the effectiveness of current cybersecurity "best practices" limited, but also they enable and encourage activities inimical to privacy. Their root paradigm is a flawed reactive one appropriately described as "penetrate and patch". Vigorous promotion encourages reliance on these flimsy best practices as a primary defense for private information. Furthermore, this paradigm is increasingly used to justify needlessly intrusive monitoring and surveillance of private information. But even worse in the long term, this misplaced reliance stifles introduction of proven and mature technology that can dramatically reduce the cyber risks to privacy.

Threat of software subversion is dire risk to privacy

Today much of the private information in the world is stored on a computer somewhere. With Internet connectivity nearly ubiquitous, it is the exception – rather than the rule – for such computers to be physically/electrically isolated, i.e., separated by an "air gap". So, protection for privacy is no better that the cybersecurity best practices defenses employed, and their evident weakness attracts cybercriminals. Billions of dollars of damage occur each year, including identity theft with massive exposure of personal data. Clearly weak cybersecurity defenses create a serious risk to privacy.

Juan Caballero's article in this issue notes that "At the core of most cybercrime operations is the attacker's ability to install malware on Internet-connected computers without the owner's informed consent." U.S. Navy research demonstrates that an artifice of six lines of code can lay bare control of a commercial operating system. The Stuxnet, DuQu and Flame software subversions have recently been detailed, and a senior researcher wrote, "Put simply, attacks like these work." I made the point myself in a 1979 article on "Computer Security: the Achilles' heel of the electronic Air Force?" where I characterized subversion as the technique of choice for professional attackers.

Best practices are not well aligned with the threat

The general response seems primarily to be a concerted push for the use of best practices, with a heavy emphasis on monitoring techniques like antivirus products and intrusion detection. For example several Silicon Valley luminaries recently presented a program with an explicit goal "To promote the use of best practices for providing security assurance". In the litigious U.S. there have even been legislative proposals to reward those who use best practices with "immunity" to lawsuits.

Yet this fails to align with the software subversion threat. A major antivirus vendor recently said, "The truth is, consumer-grade antivirus products can't protect against targeted malware." A FBI senior recently concluded that the status quo is "unsustainable in that you never get ahead, never become secure, never have a reasonable expectation of privacy or security". Similarly, an IBM keynote presenter said, "As new threats arise, we put new products in place. This is an arms race we cannot win."

But, it is even more insidious that governments use the infirm protection of best practices as an excuse for overreaching surveillance to capture and disseminate identifiable information without a willing and knowing grant of access. They falsely imply that only increased surveillance is effective. In fact, dealing with software subversion by a determined competent adversary is more intractable than scanning a lot of Internet traffic, as Flame and StuxNet amply demonstrate.

Proven verifiable protection languishes

In contrast, the security kernel is a proven and mature technology developed in the 1970s and 1980s. Rather than reactive, security is designed in to be "effective against most internal attacks – including some that many designers never considered". The technology was successfully applied to a number of military and commercial trusted computer platforms, primarily in North America and Europe. It was my privilege to lead some of the best minds in the world systematically codifying this experience as the "Class A1" verifiable protection in the Trusted Computer System Evaluation Criteria (TCSEC). An equivalent technical standard promulgated in Europe was known as ITSEC F-B3, E6.

Although no security is perfect, this criterion was distinguished by "substantially dealing with the problems of subversion of security mechanism". In other words, a powerful system-level solution aligned with the threat in just the way glaringly missing from current cybersecurity best practices. Unfortunately, at that time addressing this threat was not a market priority.

Although still commercially available, the technology has fallen into disuse in the face of the expedience of the reactive paradigm. It is particularly disappointing that now at the very time ubiquitous Internet connectivity makes privacy really, really interesting, educators and industry leaders have mostly stopped even teaching that it's possible. But today's researchers have one of those rare tipping point opportunities to lead the way to recovery from the clear and present danger to privacy by aggressively leveraging that proven "Class A1" security kernel technology.

Roger R. Schell

RESEARCH AND INNOVATION

This section features news about research activities and innovative developments from European research institutes

EVENTS, BOOKS, IN BRIEF

The "SHOWN" Platform - Safeguarding Architectural Heritage Using Wireless Networks

by Paolo Barsocchi and Maria Girardi

Regular surveys of the structure of World Heritage buildings are essential to ensure their conservation. The aging of building materials, long-term ground subsidence, and environmental vibrations are all possible causes of the deterioration of old constructions. The potentially disastrous effects of severe seismic events are an additional factor teaching us the importance of prevention. When the constraints of architectural conservation are very strict, prevention mainly means monitoring.

While great efforts have been made to monitor the surfaces of ancient monuments, the level of the technology applied to control their "structural health" is still generally quite low. A typical programme of structural monitoring simply involves technicians using classical optical instruments periodically carrying out a series of measurements. More recently, procedures have been developed to test the structural health of ancient buildings by analysing their dynamic response to natural or artificial vibrations. Such procedures are recognized as a good way to test the state of conservation of a building, and are also important aids in identifying when interventions are necessary. They consist in taking regular measurements via wired acceleration and displacement sensors, which are removed after usage. Lately, microwave interferometry technology has been used to measure the vibration of buildings at great distances. Unfortunately, these techniques generally involve high maintenance costs and thus make the continuous acquisition of the large amounts of data necessary for effective monitoring impossible. In addition, wired sensors are too invasive and aesthetically unacceptable for widespread application to the architectural heritage.

In this context, Wireless Sensor Network (WSN) technology can make an important contribution by providing an economical and relatively non-invasive instrument for real-time structural monitoring of the well-being of buildings and monuments. We envisage their employment in a not too distant future in the monitoring of entire historic areas on a large-scale, facilitating the management of maintenance operations and prompt interventions in the case of an emergency. The installation of a WSN during the construction or the restoration of a building and its connection to a central database could become an ordinary or even compulsory operation. For these reasons, the structural monitoring of ancient buildings using wireless sensor network technology seems very promising. It could provide continuous observation and real-time feedback, allowing engineers to reconfigure the network, as necessary. WSN-based technology also offers the added benefits of low visual impact and low maintenance costs.

Up to now, in the field of cultural and architectural heritage, WSNs have been mainly used to monitor large archaeological areas or some interior parameters of ancient buildings and museums, such as moisture, temperature and fire. Applications of WSN technology to the structural monitoring of ancient buildings so far have been limited to few example applications and research projects. At the moment, several issues remain to be investigated, such as the number and location of the sensors used to acquire data and the maximisation of WSN lifetime. While energy consumption is a well-known problem in WSN applications, the optimisation of the number and location of sensors is a new challenge in these technologies, which typically involve a large number of redundant sensors.

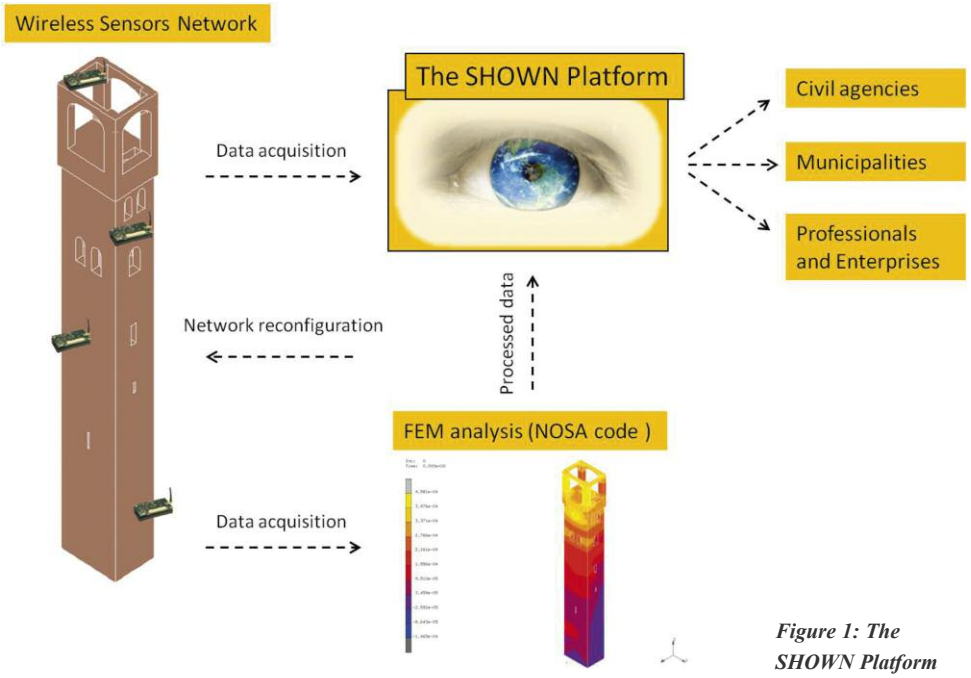


Figure 1: The SHOWN Platform

The SHOWN platform, depicted in Figure 1, will move in this direction. The research is conducted by the Mechanics of Materials and Structures Laboratory and the Wireless Sensor Networks Laboratory of ISTI–CNR and a recent breakthrough is the development of wireless networks able to communicate through a minimum number of sensors, taking into account both structural and architectonic constraints on the one hand and radio signal propagation constraints on the other. Considerable attention is also being given to energy optimisation of the network, with particular regard to the high–frequency sampling sensors, such as accelerometers.

All structural parameters are set and verified using the NOSA-ITACA code, a finite element software developed at ISTI-CNR to simulate the static and dynamic behaviour of ancient masonry.

Links:
[1] The Mechanics of Materials and Structures Laboratory: http://www.isti.cnr.it/research/unit.php?unit=MMS

[2] The Wireless Networks Laboratory (WN Lab) : http://www.isti.cnr.it/research/unit.php?unit=WN

[3] http://www.nosaitaca.it

Please contact:
Paolo Barsocchi and Maria Girardi, ISTI-CNR, Italy
E-mail: paolo.barsocchi@isti.cnr.it, maria.girardi@isti.cnr.it

Cite-as-you-Write

by Kris Jack, Maurizio Sambati, Fabrizio Silvestri, Salvatore Trani, Rossano Venturini

When starting a new research activity, it is essential to study related work. Traditional search engines and dedicated social networks are generally used to search for relevant literature. Current technologies rely on keyword based searches which, however, do not provide the support of a wider context. Cite-as-you-write aims to simplify and shorten this exploratory task: given a verbose description of the problem to be investigated, the system automatically recommends related papers/citations.

Recommender systems for scientific papers have received much attention during the last decade and some novel approaches have been experimented. We propose an innovative contextual search engine tool to address this problem. The tool exploits several aspects of the scientific literature ecosystem including an intriguing social angle derived from data provided by the popular Mendely platform. The objective is to provide pointers to existing literature given a description of the study the researcher is undertaking.

We began by building a baseline method to retrieve literature related to a fragment of text describing the research concept. This consisted of a system taking as input a textual description of the research idea and returning the most similar (according to a similarity measure) papers in the literature. To evaluate the similarity between the query and documents the "cosine similarity" metric was used. This metric calculates how many terms are in common between the query and each document and weights the terms (in query and documents) by an importance score.

We subsequently refined the baseline strategy by adopting a "Learning to Rank" approach. Within this approach the similarity between queries and documents is computed via a metric that is "learned" from samples input to a machine learning algorithm. The main difference between a search engine and Cite-as-you-write consists in how the queries are formulated: search engines are usually optimized to answer keyword-based queries, our system extracts a context from a long description of the research problem provided by the scientist.

The system consists of three modules:
- Crawler, which builds and maintains the repository of scientific literature.
- Indexer, which processes the scientific papers and builds an index on their most representative terms.
- Query processor, a specialized search engine with an advanced ranking algorithm designed for the task of retrieving related work from a detailed context.

The data used to build and evaluate our system consists of about 500 thousand computer science papers including their citations. The data was kindly provided to us by Mendeley.

The index represented under the form of an inverted index contains only the most representative terms for each paper. This trade-off in coverage, keeps down the size of the index. Fortunately. our experiments show that the loss due to reduced coverage is limited, as scientific publications usually focus on a few specific topics represented by a small number of important terms.

Following the typical approach of learning-to-rank-based retrieval systems, the ranking phase consists of two steps:
1. A cosine similarity ranking based on the title and the abstracts of the papers (ie, the baseline method)
2. A ranking function that combines all the features we have collected from the data

The second step works by adopting a technique known as similarity learning, which consists in exploiting sample data to find correlations between a set of features and a target variable. The learning method adopted is a Random Forest ensemble which uses our features based on text similarity, paper publication date, the citation network (i.e. PageRank and absolute number of citations) and the Mendeley environment (ie popularity of papers in user libraries with some normalizations: the importance of the user, defined in the social network as the number of links, and by the number of items in the library, reducing the weight when a user has lots of papers in their library). Random Forest is a very simple but powerful machine learning tool which builds several decision trees from various random samplings of the data. The result is the average of the results returned by each tree. The same strategy as that adopted in democratic voting systems.

Experiments shown in Figure 1 show the improvements on the test data over the baseline system with variations in the number of trees (x-axis).

We used the normalized discounted accumulative gain (nDCG) evaluation metric to measure the effectiveness of a ranking, ie how near relevant documents were to the top of the result lists [1]. The baseline considered (the black line) is the cosine similarity based ranking. The improvement provided by our tool on the test set with respect to the base line is summarized in Figure 2.