Searching for Experts on the Internet

Antonio DA SILVA; Bruno MANNINA; Luc QUONIAM; Hervé ROSTAING

Antonio DA SILVA, Snecma Moteurs, Groupe Snecma
Bruno MANNINA, Intelligence Process
Luc QUONIAM, Toulon / Saint -Raphaël University
Hervé ROSTAING, CRRM

Abstract

In this paper the authors describe a new method of identifying experts in a specific domain using the New Information and Communication Technologies (NICT). This method uses the technology of certain intelligent agents capable of running through a set of Web pages starting from an initial page, and collecting the e-mail addresses indicated therein. The authors then describe the method they used to set up a contacts database from these elements.

Keywords

Competitive intelligence; Expert; Internet; Watch; Knowledge networks, Contact networks; Intelligent Agents; Electronic mail addresses.

Introduction

Over the last few years, the industrial world has become truly aware of the need to optimize the use the information flows surrounding it in order to better know and understand the sphere in which it operates. In this context, the "watches", and more broadly speaking, Competitive Intelligence **[MAR 94]**, enables all the available information sources to be exploited rationally to provide an aid in strategic decision making **[CLA 99] [DOU 95]**.

This is why the notion of Competitive Intelligence involves going beyond the partial actions designated by terms such as technological watch, competitive watch, marketing watch, etc. Competitive intelligence is in fact the logical extension of the different watches, with two additional components, that bring into play the notion of networks. The first is the coordination of players from different backgrounds: salesmen, experts, decision-makers, institutions, external organizations - in short, all the people that are likely to collect informal information. The second component concerns practices of influence **[MEY 98]**, such as lobbying.

The development of the Internet, a true network of networks, has exacerbated the very notion of the network which modifies the organizational structure of companies. Indeed, prior to this the players in the company evolved in a pyramid-type configuration within a closed environment. The people in a given professional category (commercial staff, directors) developed this type of network. Today, with the multiplication of NICTs, this archaic type of organization is brought into question. Thus, the members of a company have contacts with the outside people from diverse backgrounds. This new notion of communication is drastically and fundamentally changing the structure of information sharing.

The use of NICTs to create networks of researchers to monitor evolving knowledge remains an essential element in the knowledge network. A knowledge network is a gateway between the two worlds of R&D (Research and Development) and production. It enables the evolving knowledge discovered by research work to be transferred to the production stage for industrial application **[ACH 98]**.

The interaction of internal knowledge networks is in actual fact the first stage in the Competitive Intelligence approach, because it is through these networks that the information flows will be captured, forwarded to the appropriate person for analysis, and then exploited with a view to making a decision.

However, the information held by all these networks alone does not cover the entire sphere of observation of a company. These internal networks must open up to the exterior in order to have a better approach to the company environment. Companies' experts need to update their knowledge continually to ensure that they are always at the leading edge of their field.

An expert (or analyzer) is someone who knows his field inside out. The expert analyzes and validates the information gathered by the observers, then takes part in working parties integrating transverse skills (R&D, patent rights, production, marketing, etc.) **[JAK 92]**. The collaboration between experts in different fields results in a high-added-value synthesis, a decision-making aid that is transmitted to the decision-makers for use.

The information needs of certain experts can be satisfied through internal watch networks. They must nevertheless keep an eye on the exterior. To do this it is very useful for them to have regular contacts with the world experts in their specialist field. The latter must be in regular contact with the customers, suppliers, public or private bodies and other trade associations. These special contacts will be a major asset for the company.

A participative method

Today the Internet is one of the largest sources of information in the world: fifty million computers are connected to it, that is to say fifty million computers that can be considered to be databases and informal sources of information. Faced with this enormous volume of information, when looking for information it is vital not to diverge from the subject in question, but at the same time remain sufficiently exhaustive **[SAM 98]**. At the present time, certain tools enabling users to travel on the information highways must be considered absolutely essential to achieve this end. These tools are called Intelligent Agents, or more commonly Robots **[JEN 98]**.

Intelligent agents are programs that are capable of interacting with an environment, adapting to the circumstances, making decisions or enriching their own behavior on the basis of observations they make.

The term "Agent" is currently used to designate a Robot or software capable of helping a user perform repetitive tasks. In everyday language, the term Agent is replaced by the term Robot. The vocabulary is nevertheless extremely varied, since in the literature one finds several other terms with the same meaning as Agent: Robot, Spider, Wanderer, Web Worm, Web Walker, Search Engine, Bots, Brokers.

As regards their usefulness, intelligent agents enable the user to retrieve information from the Internet, solve the problem of the unmanageable quantity of information with which the user is confronted, find information that is not indexed (not present in conventional servers) in order to have a competitive advantage, save time in the polling of servers, collect information automatically, engage a communication with the user or even other intelligent agents, find out a person's habits, or perform tasks when the user is not present.

In concrete terms, the vocation of intelligent agents differs depending on whether they are used by private individuals or companies. For **private individuals**, they guide them on the Net towards the information they are looking for, towards products at the best price at the moment of the request (journey, vehicle, etc.), help them with the management of their own documents if the agent is configured to search for information on the person's own terminal. For **companies**, they provide the same applications as for the private individual but in addition enable them to set up a technological and competition watch tool: the Web is teeming with small companies and laboratories that make known their research work and their products. They can also be used to set up a systematic watch tool: by keeping intelligent agents constantly researching into a competitor (with its name, Website address, etc.), none of the economic and strategic movements will escape the notice of the user of the said agents.

AURESYS [1] (AUtomated REsearch SYStem) is an intelligent agent resulting from the CRRM technology, which enables its users to run over all the Information Highways by looking for servers that contain the desired keywords. This method of searching for information saves time and money. A manual search through the same quantity of information would be extremely tedious. Moreover, one must not overlook the creation of lists of experts on given subjects via discussion lists and the e-mail addresses conveyed over the information highways.

AURESYS allows the creation of databases that can be queried from a distance by several users. The information is retrieved from the Internet then processed for storage in these bases. Each database responds to a personalized request from a user. Thanks to the databases created, AURESYS gives users the means of having a maximum amount of information relating to a given subject.

The user can create a corpus that can be analyzed directly by bibliometric tools, as well as General Information Files (DGI) or Strategic Information Files (DIS) **[MAN 97]**.

---

[1]

AURESYS is an intelligent agent which is a precursor of WebProcess. It was developed by Bruno Mannina as part of his PhD at the CRRM (French Scientific Center specializing in Technology Watch and Competitive Intelligence). Use of AURESYS was validated by Information Science specialists but its ergonomics and complexity made it unsuitable for the general public. This is why AURESYS was turned into another type of tool for easier operation.

WebProcess (http://www.intelligence-process.com/webprocess or http://www.searchprocess.com/eng/home.shtml for an English version) was created by Intelligence Process, a start-up set up in September 1999 by five 'information and network science' research workers including Mr. Mannina. The five scientists specialize in intelligent agents and dedicate all their efforts to the development of competitive intelligence tools using the Internet. They are established at Saint Victoret near Marseilles and develop products designed for the general public but also to meet the competitive intelligence needs of large companies.

AURESYS welcome page

AURESYS has the following characteristics:

- It complies with the robot exclusion protocol. Only authorized directories are visited.

- The databases created are updatable, hence the advantage of the system for Technological Watch.

- It offers the choice of the domain in which the robot will run its search, and the depth of search (hypertext links).

- It offers the possibility of starting the search from any page of a site.

- Thanks to the incrementation of IP numbers, AURESYS can find sites that are not registered by the conventional WEB utilities (AltaVista, Yahoo, Lycos, Search, etc.), as most other robots simply poll sites that are already indexed.

This tool can be used by any person wishing to obtain information on a given subject with continual information updating (e.g. information specialists, experts in a specific field).

In the framework of the method described above, the aim of using the AURESYS robot is to **set up networks of experts on given themes via the e-mail addresses given on certain Internet sites**. Company experts will thus be able to have access to databases of personal contacts relating to their activity.

The creation of such a database involves a number of obligatory phases.

A fundamental initial phase is a prior interview with the expert in the field to monitor. The expert must clearly define the extent of the field to monitor in order to identify the Internet site to be studied.

When a professional site is identified, pertinent results can be obtained, since generally speaking these sites are of better quality than commercial sites.

Once the site has been identified, the AURESYS E-MAILS module is parameterized to collect all the e-mail addresses on that Internet site. The results are presented as a list of e-mail addresses with the frequency of occurrence on the site. This indicator is proportional to the expert's level of activity on the monitored site.

Particular attention must be paid to the results provided by the AURESYS E-MAILS module. The e-mail addresses thus obtained are analyzed and classified per field. This type of sectorization enables the general leanings of the members of the list to be defined and individuals to be divided up according to their body of attachment: non-profit-making organizations, trading companies, universities, government institutions, military or other anonymous addresses. This sorting phase is essential for the writing of the first contact message. This message will vary depending on type of organization in which the contact works. Indeed, one does not address an academic in the same way as a soldier.
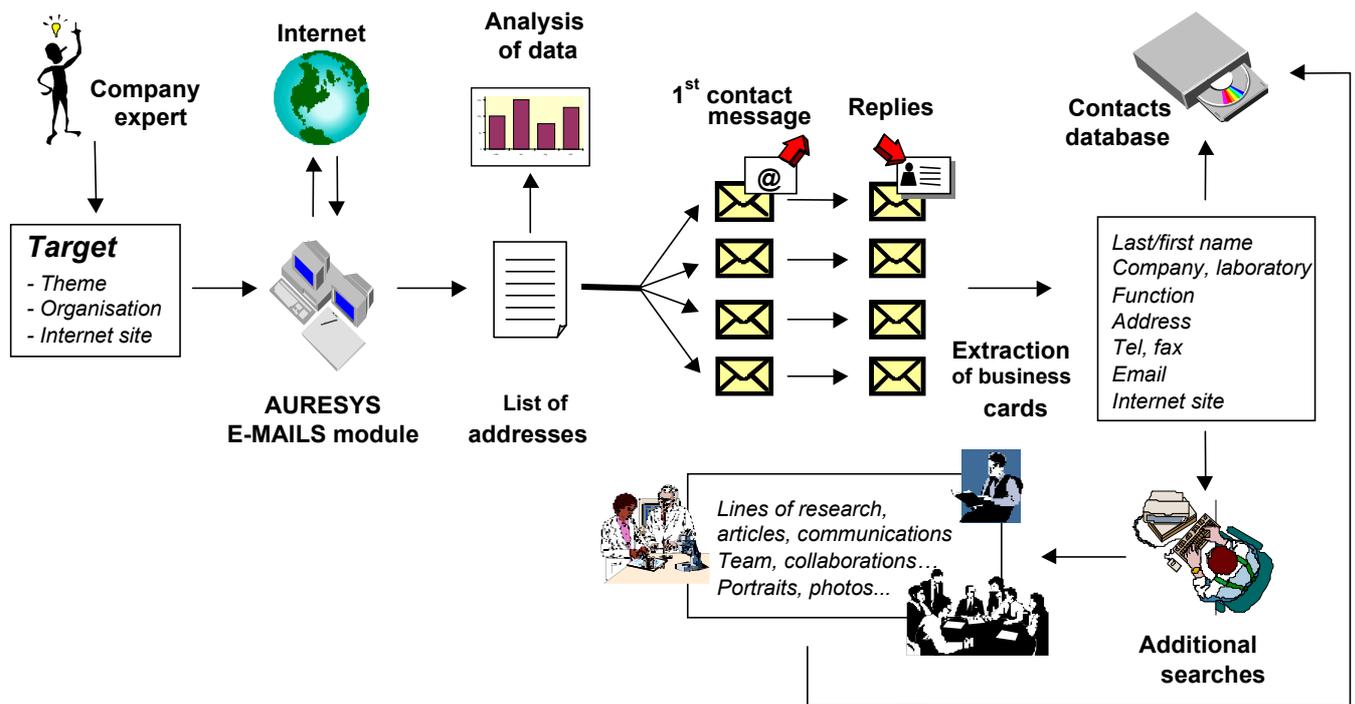
Searching for Experts on the Internet

The message must comply with certain rules of professional ethics. Firstly the company for whom one is acting must be presented. Then the true purpose of making this contact must be clearly stated: provide the experts in the company for whom this study is being carried out with a directory of external contacts with expert knowledge in the fields of activity of that company. It is also important to guarantee contacts the possibility of having access to and modifying the information concerning them.

These measures aim at protecting the private and information processing freedom of these potential future informers **[COU 81] [LOI 78]**. Moreover, they give credibility to this approach, thereby ensuring better forthcoming relationships.

The various initial-contact messages are then sent as a grouped "mailshot", while at the same time being personalized according to the attachment sites (university sites, private company sites, non-profit-making organizations, military sites, others). One must not expect a high response rate. It is usually between 20 and 25%. This can be explained by the high workload of these experts and the large number of e-mail messages they receive each day.

The next stage consists in extracting the business cards of the experts who accept to take part in this approach and saving them in the contacts database. The company experts thus have an updated electronic directory of international experts in their field of activity. Subsequently it is possible to undertake additional searches on the work of these contacts and thereby have supplementary information: lines of research, publications, research team, collaborations. A photograph of the contact is a useful element that should not be forgetten. It will, for example, help the company experts identify the person at a conference.

Application of the method to the Internet site of «Association X»

Let us now examine what the described method can give when applied to the Internet site of a trade association.

> **_NB_**_: It is important to point out that the data mentioned in this example have been intentionally masked for reasons of confidentiality. The names of the people and the organizations or companies to which they belong are not their real names. Only the proportions concerning the nature of the replies have been kept._

Association X is a trade association whose mission is to collect, analyze and distribute technical information on materials. Its operation is based on a global information-sharing network, whose members intervene during meetings and forums, and through training programs or publications. The Internet site of the Association X presents its activities and members, the training programs, the conferences and the services it provides.

This site was selected because its pages contain numerous international, university, industrial, public or private researchers. It is therefore favorable to the use of the AURESYS «E-MAILS» module which groups together all the e-mail addresses present on the given site.

Below is the configuration panel of this module.

AURESYS E-MAIL module

The search is run after setting the parameters of the AURESYS «E-MAILS» module. The results are obtained as a listing of addresses preceded by the number of pages in which they appear.

Example of list compiled by the AURESYS «E-mails» module:

Occurrence	E-mails
8	cust-srv@association.x.org
7	Person.1@association.x.org
6	Person.2@association.x.org
6	Person.3@association.x.org
6	Person.4@association.x.org
5	Person.5@association.x.org
5	Person.6@association.x.org
5	Person.7@association.x.org
4	Person.8@association.x.org
4	Person.9@association.x.org
4	Person.10@association.x.org
3	Person.11@association.x.org
3	Person.12@association.x.org
3	Person.13@association.x.org
3	Person.14@association.x.org
3	Person.15@association.x.org
3	Person.16@association.x.org
3	Person.17@association.x.org
2	Person.18@association.x.org
2	Person.19@association.x.org
2	Person.20@association.x.org
2	mem-serv@association.x.org
2	Person.21@association.x.org
2	Person.22@association.x.org
2	support@Company.A.com
1	Person.23@Company.B.com
1	Person.24@Company.C.in
1	Person.25@association.x.org
1	Person.26@association.x.org
1	Person.27@University.D.edu
1	Person.28@Company.E.com
1	Person.29@Organization.F.gov
1	Person.30@Organization.G.gov
1	Person.31@Company.H.com
1	Person.32@association.x.org
1	Person.33@association.x.org
1	Person.34@association.x.org
1	Person.35@University.I.edu
1	Person.36@msn.com
1	Person.37@Company.K.net
1	Person.38@association.x.org
1	Person.39@association.x.org
1	Person.40@association.x.org
1	Person.41@association.x.org
1	Person.42@aol.com
1	Person.43@Company.L.fi
1	Person.44@Organization.M.gov
1	Person.45@University.N.edu
1	Person.46@association.x.org
1	Person.47@association.x.org
1	Person.48@Organization.O.gov
1	Person.49@Company.P.cm
1	Person.50@Organization.Q.gov
1	Person.51@Company.R.mx
1	Person.52@Company.S.ca
1	Person.53@Company.T.com
1	Person.54@association.x.org
1	Person.55@University.U.jp
1	Person.56@Military.V.mil
1	Person.57@University.W.edu
1	Person.58@aol.com
1	Person.59@University.X.edu
1	Person.60@University.Y.ch
1	Person.61@association.x.org
1	Person.62@Organization.Z.gov
1	Person.63@Organization.A1.gov
1	Person.64@Company.B1.com
1	Person.65@Military.C1.mil
1	Person.66@association.x.org
1	Person.67@association.x.org
1	Person.68@University.D1.edu
1	Person.69@University.E1.edu
1	Person.70@association.x.org
1	Person.71@association.x.org
1	Person.72@Military.F1.mil
1	Person.73@University.G1.edu
1	Person.74@University.H1.edu
1	Person.75@Organization.I1.gov
1	Person.76@wee.com
1	Person.77@wee.com
1	Person.78@association.x.org
1	Person.79@Company.K1.net.cn
1	Person.80@University.E1.edu
1	Person.81@compuserve.com
1	support@association.x.org
1	Person.82@Company.L1.se
1	Person.83@association.x.org
1	Person.84@Organization.A1.gov
1	Person.85@aol.com
1	Person.86@aol.com
1	Person.87@Company.N1.com
1	Person.88@association.x.org
1	Person.89@University.O1.edu
1	Person.90@Military.V.mil

END OF PROCESSING.

Analysis of the addresses obtained

The addresses obtained can be classified per domain.

Domain	Frequency	Percentage	Description of the domain
.org	45	47.8	Non profit-making association
.com	7	7.4	Commercial site
	8	8.5	Anonymous address (aol, msn, etc.)
.edu	11	11.7	University
.gov	9	9.5	Government association
.mil	4	4.2	Military site
.net	2	2.1	Site dedicated to the Internet
.ca	1	1.1	Canada
.ch	1	1.1	Switzerland
.cm	1	1.1	Cameroon
.fi	1	1.1	Finland
.in	1	1.1	India
.jp	1	1.1	Japan
.se	1	1.1	Sweden
.mx	1	1.1	Mexico
TOTAL	94	100	

It can be seen that the majority of the members on this list belong to non-profit making associations. But on closer examination one sees that in fact they all belong to the same association, Association X. This is quite understandable since our study concerns that association's web site.

Other addresses belong to trading companies (.com), and can equally well correspond to private industrialists as to anonymous addresses (aol.com, compuserve.com, msn.com, wee.com). The remaining addresses are made up of universities, (.edu), government bodies (.gov), military (.mil), and various countries.

After a quick analysis, an initial contact e-mail is sent to each person. This is followed by a waiting period which can range from a few hours for the fastest replies to a few days for the slowest ones.

Analysis of the replies

The replies can be divided into three distinct categories:

First category: The replies are in fact a return-to-sender-message because the addresses in question are no longer valid.

Second category: The people have received the message but are not responding.

Third category: The people reply positively or negatively to the message sent. The positive replies include the business card of the person (surname and first name, laboratory or company, function, address and telephone/fax number, sometimes the Internet site).

Invalid addresses	No reply	Replies 28,7 % divided into:	
26.6%	44.7%	Positive 22.9%	Negative 5.8%

The majority of the return-to-sender replies concern addresses specific to the Association X. This is perhaps due to the fact that these addresses are assigned to people at conferences or meetings. Once the events are finished, these addresses are not always kept up to date.

Among the people who do not reply, it is notable that there are no replies from military bodies. In addition, commercial sites generally give a low reply rate.

As concerns the people who respond to the message, most of them give their business card. This being said, it is nevertheless found that one message is not enough insofar as these people no doubt receive a large number of e-mails each day. To establish initial contact, it would be wise to send a chase-up message to draw the persons attention to this particular message among the numerous other e-mails received.

Examples of business cards extracted from the replies obtained:

Person 60 Professor Laboratory of mechanical metallurgy Department of materials Address : aaaa aaaa aaaa aaaaa Tél : xxxx xxxx (secretariat xx1) Fax : yyy yyyyy Person.60@University.Y.ch http://www.xxxxx.html (laboratory) http://www.yyyyy.html (department) http://www.zzzzz.html (university)	**Person 24** Doctor Head mechanical metallurgy division Center for C Research Address : aaaaa aaaaa aaaaa aaa Fax : yyy yyyyy Person.24@Company.C.in
Person 55 Professor Department of Materials Science and Engineering Address : aaaaaa aaa aaaaa aaa Tél : xxxx xxxxx Fax : yyyy yyyy Person.55@University.U.jp	**Person 1** Materials Branch Address : aaaaa aaa aaaa Tél : xxxxx xxxxx, Ext. xx1 Fax : yyyyy yyy Person.1@association.x.org http://www.association.x.org
Person 42 President/Chief engineer Name-of-Society Address : aaaaa aaa aaa tel : xxxx xxxx fax : yyy yyyyy Person.42@aol.com http://www.Name-of-society.com	**Person 75** Doctor Research Laboratory Center Address : aaaa aaa aaa aaa Tél : xxxx xxxx Fax : yyyy yyyyy Person.75@Organization.I1.gov http://www.HH.org

Conclusion

This paper aims at showing how to build up a contacts database from world experts present on a professional site via the Internet.

The following advantages are found when this approach is used:

✓ Real-time updating of electronic mail addresses (unlike printed directories),

✓ Participation in the creation of the contacts database,

✓ Better availability of the world experts (those who reply are willing to take part in this type of approach),

✓ A promising level of positive replies (which leads us to believe that the experts are open to this type of approach),

✓ A new method of organizing knowledge networks.

Lastly, we would point out that this approach must be carried out in full compliance with the laws governing automated processing of data relating to private individuals. Remaining within legal framework guarantees that sound links will be created between the world experts and company experts.

Bibliography

[ACH 98] : Pierre ACHARD, Jean-Pierre BERNAT, L'intelligence économique : mode d'emploi, ADBS éditions, 1998

[CLA 99] : John L. CLARKE, Why Good People Make Bad-Decisions---Despite Great Intelligence, news & publications – Competitive Intelligence Magazine October – December 1999 (Vol. 2, N° 4), SCIP, 1999

[COU 81] : Council of Europe, European Treaties ETS N° 108, Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel

[DOU 95] : Henri DOU, Veille Technologique et Compétitivité, Dunod, 1995

[JAK 92] : François JAKOBIAK, Exemples commentés de veille technologique, Editions d'organisations, 1992

[JEN 98] : Nicolas R. JENNINGS, Michael J. WOOLDRIDGE eds., Agent technology: foundations applications and markets, Springer, 1998

[LOI 78] : Law No.78-17 of January 6, 1978 relating to data processing, files and liberties

[MAN 97] : Bruno MANNINA, Henri DOU, Bruno FLORENCE, Internet et l'innovation régionale : Auresys, le robot régional, IDT, 1997

[MAR 94] : Henri MARTRE, Intelligence économique et stratégie des entreprises, Commissariat général au Plan, La Documentation française, 1994

[MEY 98] : Roland MEYER, Intelligence économique et nouvelles pratiques sociales, Revue d'intelligence économique n°3, p 86-89,1998

[SAM 98] : Henri SAMIER, Victor SANDOVAL, La recherche intelligente sur l'internet : outils et méthodes, Éditions Hermès, 1998

Authors

About Antonio DA SILVA

Antonio DA SILVA is now preparing a Ph.D. thesis in Information and Communication Sciences at the CRRM laboratory of Aix-Marseille University in France. He is currently working on this project at Snecma Moteurs, an affiliate of Snecma Group, a major French Aerospace Engine Manufacturer. His mission is to develop a network of experts to help specialists become more closely associated in the competitive watch process.

Snecma Moteurs
✉ Site de Villaroche   Tel: 33 (0)1 60 59 86 21
77556 Moissy-Cramayel   Fax: 33 (0)1 60 59 99 35
FRANCE   e-mail: antonio.lopes-da-silva@snecma.fr

About Bruno MANNINA

Bruno MANNINA is a software R&D manager at Intelligence Process (http://www.intelligence-process.com) and a computer engineer. He is about to complete his Ph.D. thesis in Information and Communication Sciences at the CRRM laboratory. His research work covers 'an automatic analysis of informal information needed in making decisions: the case of informal information from the Internet'.

Intelligence Process
✉ Boulevard Abbadie Espace Rossignol,   Tel: 33 (0)4 42 77 52 22
Bâtiment 2   Fax: 33 (0)4 42 77 52 23
13730 SAINT-VICTORET   e-mail: bruno.mannina@intelligence-
FRANCE   process.com

About Professor Luc Quoniam

Professor in Information Science within 12 years, Participate to the Competitive Intelligence PhD program in Marseille France (http://crrm.u-3mrs.fr).
Research activity in Competitive Intelligence oftenly in conjunction with Institutions, Industries;more specifically in data collecting, Text Mining, Document Mining, Textual and Data Analysis.

LePont-CRRM University of Toulon
✉ BP 192   e-mail: quoniam@univ-tln.fr
F-83957-LaGarde Cedex
FRANCE

About Hervé ROSTAING

Hervé Rostaing is associate professor in CRRM's laboratory at the Aix-Marseille University, France. He is teaching in two Information Sciences degrees : Bachelor of Science (an online distance learning with Internet facilities) and Post-graduation on Competitive Intelligence and Technology Watch. His research activities is concerned by statistical textual treatments and text mining applied to Technology Watch. He is author of a French book on "Bibliometrics and its techniques".
www site: http://crrm.u-3mrs.fr

CRRM
🖃 Faculté des Sciences   Tel: 33 (0) 4 91 28 87 46
av. Escadrille Normandie-Niemen   Fax: 33 (0) 4 91 28 87 12
13397 Marseille Cedex 20   e-mail: rostaing@crrm.u-3mrs.fr
FRANCE