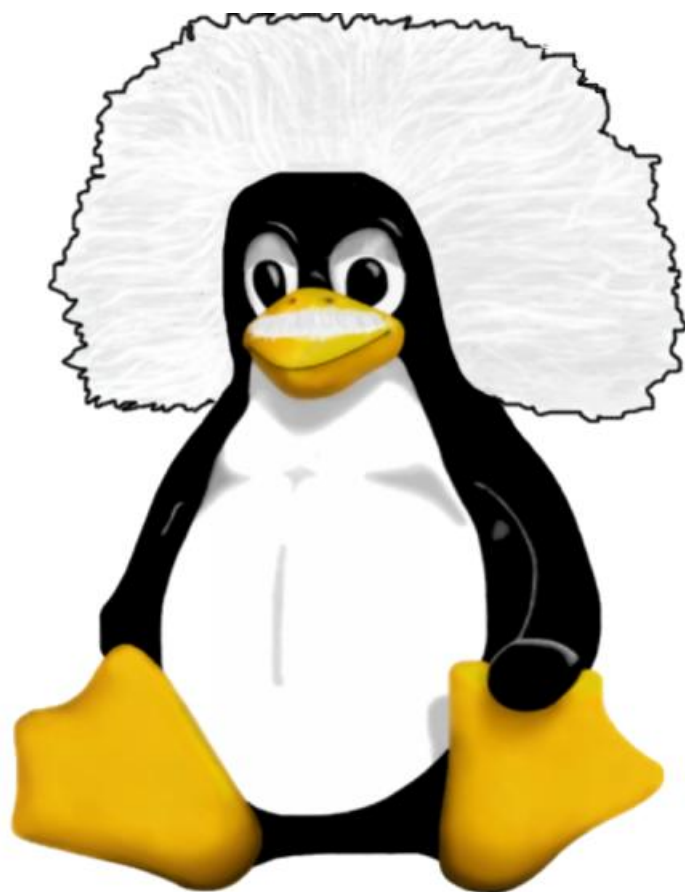


SCIENtific Searcher: A System for Scientific Information Searching and Managing



Amanda Kisuy Chi Cerrito
Antonio Téllez Flores
Francisco Federico Raggi Cardenas

Contents

Cover	1
Contents	7
I Tutorial	1
1 What is ScienSe?	3
2 Antecedents	5
3 History	7
4 Basic system's architecture	9
5 License	13
6 ScienSe's Advantages	15
7 Developers	19
7.1 Francisco Federico Raggi Cardenas	19
7.2 Amanda Kisuy Chi Cerrito	20
7.3 Antonio Téllez Flores	21
8 What is next...	23
II User's Guide	25
9 ScienSe Features	27

10 Requirements for the client's side	29
11 Using Sciense	31
11.1 Adding publications to the database	31
11.2 Adding authors	32
11.3 Adding references	33
11.4 Searching publications	34
11.5 Searching Authors	35
11.6 Searching References	36
11.7 Searching Publications in the temporary database	36
11.8 Searching authors in the temporary database	37
11.9 Adding from the automatic index	38
12 Understanding the error messages	41
13 Modifying the information of the database	45
14 Getting help and support	47
15 Contributing with the project	49
III Administrator's Guide	51
16 System Requirements	55
16.1 Hardware	56
16.2 Software	56
16.2.1 For executing Sciense	56
16.2.2 In order to see the documentation	57
16.2.3 For compiling Sciense	57
16.2.4 For modifying the documentation	58
16.2.5 For participating in the Sciense project	58
17 Preparing the environment	59
17.1 Installing the software required	60
17.2 Configuring PostgreSQL	60
17.3 Configuring Apache	61
17.4 Configuring htscience	61
17.5 Configuring Tomcat	62

<i>CONTENTS</i>	5
18 Installing ScienSe	65
19 Processes executed by ScienSe	67
20 The indexing process	69
21 The checking process	71
22 The adding information process	73
23 The searching process	75
24 Maintenance	77
25 Backing up the information	79
26 Getting support	81
IV Programmer's Guide	83
27 Basic structure	85
28 APIs	87
29 Converters	89
30 Plug-ins	91
31 Web Interface	93
32 The database	95
33 Security	97
V Developer's Guide	99
34 Details of design	103
35 Development tools	105

36 HTSciense	107
37 The database	109
38 The web application	111
39 Security details	113
40 The documentation	115
41 Future Projects	117
VI References	121
42 Java Servlets	123
43 Java Server Pages	125
44 Javascript	127
45 Tomcat	129
46 Apache	131
47 Postgresql	133
48 GNU Wget	135
49 RPM	137
VII FAQs	139
50 General	141
51 User	145

CONTENTS

7

Bibliography

155

Bibliography

155

Part I

Tutorial

Chapter 1

What is ScienSe?

ScienSe is a system of software that tries to provide to researchers, scientists, students and in general to any person who has some kind of relationship with the science, the opportunity to obtain and manipulate the information related with scientific publications stored either in the Internet or a local Intranet, in a fast, simple and easy way.

ScienSe is a project developed by members of the [Faculty of Sciences](#) at the National University of Mexico ([UNAM](#)) under the philosophy of free software, which allows a wide access to the system, its improvement and application.

Chapter 2

Antecedents

Next, we will give some subjects related with the nature of this project that will help to understand its reason of being and the motivations behind.

The scientific publications are important by 3 fundamental factors:

1. They allow that a researcher receives criticisms, to his work, from other researchers of the same area, generating a feedback process that improves the scientific work.
2. They allow that new researchers begin their work based on others already made and do not start every research from scratch.
3. They allow to evaluate the quality of the activity of scientists by means of the impact of their publications, which is moderated generally by 2 factors; first it is the publication on a magazine of great prestige in the field and the second is the number of references from other researchers to their work.

At the present time it is taken place, to call it like this, a crisis in publications of scientific character, consequence of two practically unstoppable phenomena. First it is the abusive price of subscriptions cost to the magazines and the second is the Internet irruption.

The prices of the magazines on the last 10 years have increased their prices every year between 12% and 16%. This has meant that worldwide

universities and libraries have begun to cancel titles of magazines in a massive form, because it is already to them impossible to assume these prices. This is more visible in the Third World Countries universities where the expenses in research and education tend to the loss. The problem is so, that great American and European universities have even reduced more than 21% of the contemplated expenses for the acquisition of magazines.

This problem has arrived to some scientists and professors members of scientific committees in magazines of prestige and, knowledgeable of this crazy increase of prices, have left the committees of the magazines themselves since they do not agree in returning to pay by a knowledge of which the university already was proprietor and of course, they do not accept the absurd that has arrived, in where in some cases, are the own researchers those who must pay to the magazines so that their works can be published.

In the presence of this situation, certain scientists' groups are fomenting alternative projects based on the edition of scientific electronic magazines with free or low cost with access via Internet. Even the universities or researchers' groups are returning to publish their magazines without the needed of intermediaries. It is beginning to extend the opinion that is due to take more advantage of the WWW, been born in the universities and research centers, the way to develop a new diffusion of the Science.

It is in this context that appears ScienSe, which is a Free Software that tries to contribute to solve the diffusion problem of the Science, particularly of those publications of scientific character. Free software took exactly the scientific philosophy of sharing and now the Science can take advantage of this model in order to solve a problem that worries, more everyday, to the scientific communities anywhere in the world.

Although, other aimed projects have arisen to solve this problem, ScienSe combines free software with the idea to make of free access the scientific publications and will base its success on the power of scientists' and free software programmers' collaboration.

Chapter 3

History

In April 2003, was carried out, an interview between the undergraduate students of Computer Science in the Faculty of Sciences at the UNAM: [Amanda Kisuy Chi Cerrito](#) and [Antonio Téllez Flores](#) with the [ScD Francisco Federico Raggi Cardenas](#), researcher class D in the [Mathematics Institute](#) at the [UNAM](#), in which the primary target, was to choose some of the problems that ScD. Raggi faced during his scientific performance that could be solved by a computer program, which would be developed by the students before mentioned like a social service project.

Is in this way, that exploring the different possibilities, Dr. Raggi mentions the inconvenience of, that researchers like him, cannot know who makes reference to their works, due to the great amount of time that takes doing manually the search of that information, and this originates that in most of the cases, the real proportions at the research works, get lost a little compared with those people that can take the time to make such job.

And that is the idea that gives origin to ScienSe.

With that in mind Antonio Téllez and Kisuy Chi focus to look for alternatives of solution.

The first step was to carefully make an exposition of the problem, from which be able to extract the fundamental characteristics that a software should have for being able to achieve that objective. Once obtained such requirements a field research was made in which, those software systems that

fulfilled at least one of the system requirements were looked for and analyzed.

With this, the disadvantages that cause those systems do not achieve the users' expectations could be localized. This took us to redefine some of the requirements established at the beginning with the purpose of not incurring, as far as possible, on such errors.

The following step was to make another investigation, but this time, focused to find the best already existing computational technology that could be used for the system development.

Due to this search, was more fruitful from which it was hoped, it was noticed that the system expectations could grow up if it led the maximum advantage from the characteristics of the tools that would be used, this took us to again extend the system requirements.

After this, we occurred to the task of looking for a project's name. It had to reflect, as far as possible, the nature of the product, also it had to be able to attract the public to which it was directed and finally, it had to be a name easy to remember and difficult to forget. It was with these goals that the idea to call the product "SCIENTific SEArcher" (*ScienSe*) was born, that was easy to remember, common, attractive and the sufficiently descriptive.

In this stage we were ourselves forced to suspend the project by reasons beyond our control.

It was until August 2003 that we had the opportunity to retake the project, initiating with the software implementation, once we had clearly identified the different modules of the system.

Everything elapsed as it was hoped until the middle of December when we again had to stop the project, due to a series of unexpected economic conflicts. And the project continuous delay until February 2004. Although this time we would reinstate it until culminating with the first version in May of the same year.

Chapter 4

Basic system's architecture

When the planning of the system was carried out, different alternatives to design the application were thought, some of them were the following ones:

1. A single-user system to be acceded from an only machine.
2. A multiuser system that would be acceded from an only machine.
3. A client-server system where the client is an application program.
4. A client-server system where the client is a web interface.
5. A distributed system where the client can be a web interface.

As the main idea of the system is to take the scientific information to a greater amount of people and to have a very accessible system, options 1 and 2 were discarded immediately.

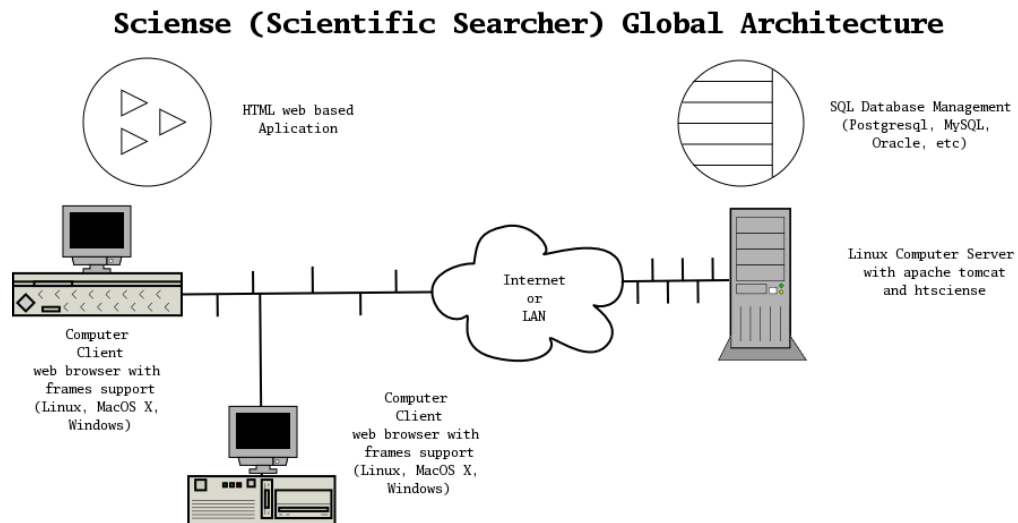
Option 3 was had in consideration, but the effort to develop clients for different platforms was too much and useless, even using highly portable programming languages, because of the installation and usability test that should be made on each one of them, which would even increase the work and the amount of necessary support, so we decided that the best thing would be a web interface based on HTML that provides similar functionality without its disadvantages.

With this on mind, we only had already left options 4 and 5. Between these it convinced to us the number 5, because a distributed system would

cause that the expenses in processing and storage diminished, in addition of that it would improve the cooperation efforts for sharing information of scientific kind. The problem is that the complexity of the system would cause that we took more on the development, which would not have to be very long taking in consideration that it was only a social service project.

So at the end we were decided by option 4, although having in mind that the system could be extended in a future to be a distributed system.

This way ScienSe was designing with the following architecture based on the client-server model:



The system is made up of 3 main parts:

1. **Indexing motor**, htscience; based on htdig (<http://www.htdig.org>), it has two important functions: the first is to index the information of the indicated sites and to store this information on a database in order to make searches, by using a cgi script, later; the second is that it looks for research articles and analyzes them in order to extract their information, that will be kept on a relational database.
2. **Relational database**; Here the information of the research articles (as title, authors, year, references, etc.) is stored, with a consistent format

and with a structure that makes the search of information easier. This database is made up from the automatically collected, by the indexing motor, data, after be checked, although the information can also be stored without any checking, in a “temporary” database.

3. **Web application;** By means of a web browser with frames support it allows to consult all the information contained on the relational database, using for it different criteria search, through any computer with access to the system.

This model was chosen because it makes the administration tasks more easier, since when all the software is only in one point, the updates are reflected immediately in the clients, besides allowing the system works easily in heterogeneous environments where different platforms of hardware and software coexist as happens in most of the universities and research institutes. Another advantages is that the system can be acceded by means of Internet or a intranet.

Furthermore, it reduces the requirements to the most, because a user can enter to the system only by using a computer, with a web browser with frames’ support, that can enter by network to the system.

Chapter 5

License

ScienSe is an open source product, protected by the [GNU General Public License](#) (GPL), which allows the use, modification and/or revision of all the software (including the source code).

GNU Copyleft, or GPL, allows to the creators of a program to conserve their legal authors' rights, but it also allows the others to copy, modify and sell the new resulting program. Nevertheless, when doing it, they cannot limit any similar right to those that obtain the new software. If they sell either the program as it is, or a modification of it, they must provide the source code.

For a greater reference consult the GNU site (<http://www.gnu.org>) and the Free Software Foundation (<http://www.fsf.org>).

Chapter 6

ScienSe's Advantages

Next we will show some of the advantages that the ScienSe system has.

Free

ScienSe is under the General Public License (GPL), it allows you or somebody else to take, review, improve or adapt the source code of the software without any special proceeding, whenever it does not fail to fulfill the GPL to that ScienSe is subject.

Gratuitous

Most of the available systems require the payment of expensive memberships so as to access them, this originates that those programs are only within reach of who wants and can pay by them. On ScienSe, we thought that the scientific information must be within reach of who requires it at any time, by this reason, we put this software to the disposition of everybody that requires without waiting for it any kind of remuneration.

Although by his license, ScienSe can be commercialized, the intention of the developers is to always count with a version without any greater cost for the user that means to obtain a copy.

Standardized

When designing ScienSe, we did it thinking about the interoperability and by this reason we based it on standards, this allows the stored information be accessible from any other program that handles the standards, with this, the user is not tied to the system once it has initiated its use, as happens with other proprietary systems; with ScienSe, if at any moment you decide to change yourself of software, you can do it without losing the valuable information that handles until that moment. We made our better effort so as to get ScienSe fulfills at least the most popular standards on every level of the system.

Automatic

Unlike other existing systems, ScienSe is more than a single information manager, since it has an integrated indexer. This means that it is not required that the data of the articles to be provided manually, ScienSe can periodically consults the sites that are indicated to it in order to obtain from them the works that have been published recently, get the outstanding data from them, organize these information and store them in a structure that makes its consultation easier.

High Technology

We used technologies like PostgreSQL, Tomcat, Java Server Pages, Java Servlets, Apache. In addition to programming languages like XML, C++, Java, Javascript. All of them with wide support and proven quality, which causes that ScienSe guarantees a good performance and an easy maintenance. For example XML is becoming the standard for the exchange of information on the World Wide Web. C++ and Java are the languages most used, either in the academic environment or in the industry. Apache, Tomcat and PostgreSQL are applications with recognized quality and some of them already have been used in academic and commercial environment with great success.

Non Elitist

Almost all the systems of this kind in the market, taking advantage of the great necessity that the users has for the information they handle, and the shortage of such, they occur the luxury to restrict, as most as they can, the access to their systems anyway.

A few are, those that can pay the membership and still less those that have the privilege of enjoying a mention within such systems, mainly on the third world countries where the research budgets has a tendency to the loss. This originates problems such as that if one has the access to one of these sites and is searching some information, one does not have a total certainty of finding it there, because each one is restricted to a few privileged sites, thus in order to increase of significant way the probability of success is necessary to count with one membership in each one of the existing systems, thing that is practically ridiculous and unnecessary. With ScienSe this it is an old problem since as you can decide the sites where to look for the information, at any time you can get that information without greater proceedings with the security of that you have access to the information, you have chosen.

Secure

The development of the system has been carried out following the best techniques in order to make a safe programming in all the levels. In addition its architecture based on layers makes the security easier to guarantee in each level.

Truthful and flexible

Due to the characteristics of ScienSe, the obtained data could be kept under two different modalities, depending this, of the process that is followed for it. The process to add new information to the database can be one of the following:

- **Automatically.** Once the indexer has finished its work, one can send the automatically obtained data to a “temporary” database without any check. The data of this category can be consulted by the user

whenever he asks for it, but with the understood of that they can have some errors or vagueness, since they were obtained by an algorithm based on probabilities. If the user checks that information later, then they become considered as part of the kept ones by the following category.

- **Reliably.** Once the indexed process has finalized, the user will be able to check one by one the data collected by the program allowing of this form to correct those erroneous fields automatically obtained and this information could be kept in a database with the information of equal origin, with which the consultations to this database offers total veracity to the user.

Efficient

Due to the ScienSe's architecture each one of the different layers has been implemented as efficient as possible, using the best algorithms in the implementation.

Friendly

The ScienSe's developers team knows that the ease use is an important factor for the success of any software. Is for that reason that has been given special attention to the final-user interface. For it, we have become attached, during the design, to the recommendations done by the creators of the most successful user interfaces of the market, and to the studies of usability.

Chapter 7

Developers

Since we have already exposed, in the history of the project, the author of the idea behind ScienSe is the ScD Francisco Raggi, whereas the design and the implementation of ScienSe have been carried out by the bachelor students Amanda Kisuy Chi Cerrito and Antonio Téllez Flores as their social service project.

7.1 Francisco Federico Raggi Cardenas

Nowadays, he is a titular researcher class D in the Mathematics Institute at the National University of Mexico (UNAM). In 1963 receives the degree of Mathematician, after that, he makes the master studies in the Harvard University. And finally he obtains the doctorate in the Faculty of Sciences at the UNAM, being the specialty of this Mathematics, presenting the titled thesis: “Las unidades en anillos de grupo” (“The units in group rings”). His area of interest in this moment is the ring theory.

Apart from his well-known passion by the mathematics, ScD Francisco Raggi has many other diverse interests. Tennis and fronton are some of his favorite sports and he practices them frequently. A great reader and admirer of Jorge Luis Borges, Francisco Raggi enjoys widely the conversations about politics just like the cinema and the nature.

7.2 Amanda Kisuy Chi Cerrito

A Computer Science student in the Faculty of Sciences at the UNAM. Her academic interests turn around the artificial intelligence, cognitive sciences, psychology, neuronal networks and computational algorithms. She has given courses of Analysis of Algorithms, Parallel Programming, Introduction to Computer Sciences and C Programming in the Faculty of Sciences at the UNAM.

At her 23 years, Kisuy is possessing of a great intelligence, perseverance, creativity and discipline. In spite of the problems and the uncertainties, because her father died when she was so young and since child she has had to work to gain the life, she always has been one of the best students of each one of the educative levels she has been, getting recognition to her academic excellence in several times. An interesting thing is the fact that in spite of had studied practically all the degree without having a computer at home, she got excellent grades.

Furthermore, she has shown a great interest about the science, participating, since an early age, in events such as: the Olympic Games of Physics, Mathematics and Chemistry, the National Mathematics Contest Pierre Fermat, the Short Stays in the Institute of Geophysical at the UNAM within the program “Young’s Towards the Research”, the Mathematics Contest I, II and III of University Initiation, the Preuniversitary Congress of Sciences, among others, getting always a good performance everytime. She was a speaker in the “V National Colloquy of Theory of Codes, Cryptography and related areas”. after developing, together Antonio Téllez Flores, a coding method based on neural networks.

She also enjoys of Linux Operating System and the C++, C, Java, Prolog and Scheme programming. Addict to the mathematical and inventiveness games.

In addition to the Science, her other interests turn around arts like dance, figure skating, music, theater, architecture and literature. And sports like martial arts, gymnastics, tennis and swimming.

7.3 Antonio Téllez Flores

A Computer Science student in the Faculty of Sciences at the UNAM. Interested in programming languages, analysis and design of algorithms, operating systems, computer architecture and cognitive sciences. He has given courses in C Programming, Analysis of Algorithms, Numerical Analysis and Introduction to Computer Science.

To his 25 years of age, Antonio is possessor of a great imagination, creativity, leadership and self-sufficiency, that has surprised in several occasions to those who know him as by the innovator of his ideas, as by his great ability for the planning, designing and leadership of software projects. In spite of his youth and thanks to his scientific qualities, he has participated in congresses always speaking about his own research works.

Widely recognized in the area, by his deep knowledge of Linux Operating System. Innate programmer that handles languages like: C, C++, Java, Fortran, Haskell, Perl, Scheme, among others, being C++ his favorite language, until the moment.

As all good scientist his knowledge are not limited to computation, because he is a great addict to history, mathematics, economy, literature, philosophy, music, cinema, physics, biology and psychology. His huge passion by history has created him great predilection by excellents Mexican historians, in addition to an addiction to the museums and archaeological zones. The economy has been for Antonio a scientific form to understand better the world.

His thought has been under the influence of authors like: Friedrich Nietzsche, Miguel de Cervantes Saavedra, Sade, Carl Jung, Sigmund Freud, Issac Asimov, Charles Darwin, Desmond Morris, A. Oparin, F. Engels, I. Kant, J. J. Rousseau, V. I. Lenin, Nicolas Maquiavelo, Carl Marx, Rene Descartes, Bertrand Russell, Juan Jose Lopez-ibor, Havelock Ellis, Marvin Minsky, and others.

Chapter 8

What is next...

In the following sections we will present documentation for those who want to know more about the ScienSe project, either as a end user, administrator, programmer or project's member, and even if you simply want to know more about ScienSe.

The following sections are:

- **User's guide:** It contains information that will be of utility to end users of ScienSe. It speaks about the requirements and gives help to use the system in a productive way.
- **Administrator's guide:** It contains the topics that will be of help to the administrator of a site that has installed or that glides to install ScienSe, it contains subjects like: installation, configuration and maintenance.
- **Programmer's guide:** Is a guide that tries to orient the programmer who wants to personalize ScienSe to his particular needs.
- **Developer's guide:** Contains indispensable information for who wants to work in the ScienSe project.
- **References:** It contains information about the technologies used in ScienSe.

- **FAQ:** It contains answers of first hand to frequently done questions about the ScienSe project.

Part II
User's Guide

Chapter 9

ScienSe Features

The main characteristics of ScienSe are:

1. It indexes documents from multiple web sites given by the user.
2. It looks for documents in several formats including: plain text, adobe PDF, Postscript, Microsoft Word, Microsoft Excel and HTML.
3. It organizes and stores the searches in a relational database that fulfills the SQL.92 standard.
4. It allows the search of the publications stored in the databases using several criteria like: name of the author, title of the publication, institution of the author, date of publication, editorial, organization, institution of publication, kind of publication, keywords or combinations of these.
5. It has a web based interface of simple use.
6. The indexing of the information is automatic and based on several heuristic to look for the publications and to store the information in the database.
7. It is free software covered by the GNU General Public License.

Chapter 10

Requirements for the client's side

The requirements for the client's side are minimum, the necessary things are the following:

1. A web browser with support of frames like: Netscape, Mozilla, Konqueror, Opera, Galeon, Epiphany, Firefox, Safari, Internet Explorer, etc.
2. A network access to the system: for instance, if the system is available on Internet is necessary to count with a Internet connection, and if it is in an Intranet, is necessary to have access to the same one.
3. It can also be desirable that the web browser counts with JavaScript and CSS 1.0 or higher support, but is not indispensable, although the great majority of the present browsers supports these characteristics.

Chapter 11

Using Sciense

11.1 Adding publications to the database

1. Go to the section called “Add to database” of the left frame.
2. Select in the pull-down menu the option: “Add Publication”.
3. If within the lower right frame does not appear the page for adding publications your browser possibly does not have Javascript support or it has it deactivated, in that case press the “Go” button that is under the pull-down menu.
4. Once it appears; in the lower righth frame, the page titled “Add Publication”, we begin to fill the data of the publication.
5. In case that you feels confused about how to filling any data, selects with the mouse the question symbol that this next to the data and a contextual aid will appear, either like an emergent window or a web page depending on the javascript support of your browser. It is necessary to point out that certain information is obligatory and without it the process of addition to the database won continue, this obligatory information is marked with an asterisk.
6. Once you have finished filling the information of the publication, press the “submit” button, if there are no errors that the system can detect, like empty spaces on an obligatory field, a year or a URL invalid, then another form will appear in where according to the type of publication that you have selected there will be requested additional data.

7. Answer the data in the same way and at the end press the “submit” button again.
8. If your browser does not have javascript support and the form he has been filled incorrectly, an error page, informing about the error, will be show.
9. Next it will appear a page that will show the new added publication by some seconds and after that a options menu will be displayed. If you does not want to wait until that, you can select the link called “Continue”.
10. If a page with the text “This Publication already exist in the database system” appears, it means that the publication you wanted to add already existed in the database.
11. This finishes the adding publication process.

11.2 Adding authors

1. Go to the section called “Add to database”, of the left frame.
2. Select in the pull-down menu the option: “Add Author”.
3. If within the lower right frame does not appear the page for adding authors your browser possibly does not have Javascript support or it has it deactivated, in that case press the “Go” button that is under the pull-down menu.
4. Once it appears; in the lower right frame, the page titled “Add Author”, we begin to fill the data of the author.
5. In case that you feel confused about how to filling any data, selects with the mouse the question symbol that this next to the data and a contextual help will appear, either like an emergent window or a web page depending on the javascript support of your browser. It is necessary to point out that certain information is obligatory and without it the process of addition to the database won continue, this obligatory information is marked with an asterisk.

6. Once you have finished filling the information of the author, press the “submit” button, if there are no errors that the system can detect, like empty spaces on an obligatory field, an electronic mail or a URL invalid, then, in another page the author’s information will be display by some seconds, after that a options menu will appear. If you do not want to wait for it, you can select the link called “Continue”.
7. If your browser does not have javascript support and the form he has been filled incorrectly, an error page, informing about the error, will be show.
8. If a page with the text “This Author already exist in the database system” appears, it means that the author you wanted to add already existed in the database.
9. This finishes the adding author process.

11.3 Adding references

1. Go to the section called “Add to database” of the left frame.
2. Select in the pull-down menu, the option: “Add Cite”.
3. If within the lower right frame does not appear the page for adding authors your browser possibly does not have Javascript support or it has it deactivated, in that case press the “Go” button that is under the pull-down menu.
4. Once it appears; in the lower right frame, the page titled “Add Cite”, we begin to fill the required data.
5. In case that you feel confused about how to filling any data, selects with the mouse the question symbol that this next to the data and a contextual help will appear, either like an emergent window or a web page depending on the javascript support of your browser. It is necessary to point out that certain information is obligatory and without it the process of addition to the database won continue, this obligatory information is marked with an asterisk.
6. After filling the asked data press the “submit” button.

7. If your browser does not have javascript support and the form he has been filled incorrectly, an error page, informing about the error, will be show.
8. If a page with the text “This Cite already exist in the database system” appears, it means that the reference you wanted to add already existed in the database.
9. This finishes the adding a reference process.

11.4 Searching publications

1. Go to the section called “Internal Searching” of the left frame.
2. Select in the pull-down menu the option: “Search Publication”.
3. If within the higher right frame does not appear the page for searching the publications, your browser possibly does not have javascript support or it has it deactivated, in such case press the “Go” button that is under the pull-down menu.
4. Select in the pull-down menu, the type of publication that you want to search. By default it searches in all the types of publication.
5. Select the type of searching in the pull-down menu, by default the search is by keyword.
6. Write the text to be looked for in, the text square, this text must of being according to the selected type of search. The search is done taking the elements that begin with the text to look for. This square accepts the symbol '%' that shows the elements of the database alphabetically sorted depending on the type of search.
7. In case that you feel confused when filling any data, select with the Mouse the symbol '?' that this next to the data and will appear a contextual aid, either in form of an emergent window or of a web page depending on the javascript support of your browser.
8. Once you have finished selecting the criteria search and have written the text to be looked for, press the “search” button. You can also clean the fields by pressing the “clear” button.

9. If you did not fill the text field, either an emergent window or an error page, depending on the javascript support of your browser, will be displayed warning it to you.
10. If the text field was filled, a page with the results of the search will appear. If there were results, they will be showed five by five, and at the end of the page the searching form is shown again. Each one of the fields with the results has a button called “More Information...” that shows detailed information of the publication according to its type, within this information there are also, the number of publications that it refers and the number of publications that referred it. These results are displayed as a link that when selecting it, shows such publications.
11. This finishes the search for publications process.

11.5 Searching Authors

1. Go to the section called “Internal Searching” of the left frame.
2. Select in the pull-down menu the option: “Search Author”.
3. If within the right higher frame does not appear the page for searching authors your browser possibly does not have javascript support or it is deactivated, in that case presses the “Go!” button that is under the pull-down menu.
4. In the searching authors page there are 2 text fields to search by name and by institution, at least one of the fields will have to be non empty and if both fields contain information the search will become of such form that will fulfill both criteria.
5. In case that you feel confused about how to fill some data, selects with the mouse the question mark that this next to the data and a contextual aid will appear, either as an emergent window or a web page depending on the javascript support of your browser.
6. Once you have finished selecting the criteria search and have written the text to look for, press the “search” button. Although you can also clean to the fields by pressing the “clear” button.

7. If you did not fill at least one of the text fields either an emergent window or a error page will appear depending on the javascript support of your browser.
8. If at least one field was filled a page with the results obtained by the search will be show. If there were results they will be showed five by five. At the end of the results page the search form will be show again. Each one of the fields with the results has a button called “More Information...” that shows detailed information of the author, within this information the number of his publications, the number of publications he references and the number of publications that references his work. These results are displayed in a link that when selecting shows such publications to us.
9. This finishes the searching authors process.

11.6 Searching References

In order to see how to search references see “Searching authors” and “Searching publications”.

11.7 Searching Publications in the temporary database

1. Go to the section called “Automatic Index” of the left frame.
2. Select in the pull-down menu the option: “Search Publication”.
3. If within the right higher frame does not appear the page for searching publications, your browser possibly does not have javascript support or it is deactivated, in that case press the “Go!” button that under the pull-down menu.
4. Select the type of searching that you want in the pull-down menu, by default it looks for by keyword.
5. Write the text to be search in the field, this text must be according with the type of searching. This field accepts the ‘%’ symbol which shows

all the elements of the database sorted alphabetically, depending on the searching type.

6. In case that you feel confused about how to fill any data, selects with the mouse the question symbol that this next to the data and a contextual aid will appear, either as an emergent window or a web page depending on the javascript support of your browser.
7. Once you have finished selecting the criteria search and have written the text to look for, pressed the “search” the button. Although you can also clean the fields by pressing the “clear” button.
8. If you do not place a string in the text field either an emergent window or an error page warning to you will be display, depending on the javascript support of your browser.
9. If all the text fields were filled a page with the results of the search will be showed. If there were results they will be displayed five by five. At the end of the results page the search form is showed again. Each one of the fields with the results has a button called “More Information...” which shows detailed information of the publication according to its type. There is also, a button which allows to add this publications to the permanent database.
10. This finishes the searching publications in the temporary database process.

11.8 Searching authors in the temporary database

1. Go to the section called “Automatic Index” of the left frame.
2. Select in the pull-down menu the option: “Search Author”.
3. If within the right higher frame does not appear the page for searching authors it is possible that your browser does not have javascript support or it is deactivated, in that case press the “Go!” button that is under the pull-down menu.
4. In the searching authors page there are 2 text fields for doing the search: by name and by institution, at least one of the fields will have to be non

empty and if both fields contain information the search will become of such form that will fulfill both criteria.

5. In case that you feel confused about how to fill some data, selects with the mouse the question symbol that this next to the data this will display a contextual aid, either as an emergent window or as a web page depending on the javascript support of your browser.
6. Once you have finished selecting the criteria search and have written the text to look for, press the “search” button. You can also clean the fields by pressing the “clear” button.
7. If you did not place a text in at least one of the fields either an emergent window or a error page warning to you will be displayed, depending on the javascript support of your browser.
8. If at least one field was filled a page with the results of the search will be showed. If there were results, they will be show five by five. The search form is showed again, at the end of the results page. Each one of the fields with the results has a button called “More Information...” that shows detailed information about the author, within this information there is the number of his publications. There is also a button that allows to add this author in the permanent database.
9. This finishes the searching authors in the temporary database process.

11.9 Adding from the automatic index

1. Go to the section called “Automatic Index” of the left frame.
2. Select in the pull-down menu the option: “Add Publication”.
3. If within the right higher frame does not appear the page to look for publications it impossible that your browser does not have javascript support or it is deactivated, in that case press the “Go!” button that is under the pull-down menu.
4. Next it will appear a page for access to the system, that will ask you for a name of user and a password. It is necessary to emphasize that in this area only the system administrator must have access. In order

to know more about who can access to this section, see “The administration guide”. In this page there are 2 text fields to be filled both are obligatory.

5. In case that you feel confused about how to fill any data, selects with the mouse the question mark that is next to the data and a contextual aid will appear, either as an emergent window or a web page depending on the javascript support of your browser.
6. After filling the data press the “login” button. Or you can also clean the fields by pressing the “clear” button. It is necessary to emphasize that by security, in the password field, the symbol ‘*’ will appear instead of every character you give.
7. If you did not fill both fields, either an emergent window or an error page, will be showed warning it to you, depending on the javascript support of your browser.
8. If you introduced an incorrect user name or password an error page will be displayed informing you about the error and inviting you to try it again.
9. If you introduced correct values it will appear a page titled “Add Publication From Automatic Index”. In this and in the next pages, it appears a message that says: “Click to close this session LogOff”, this link is for turning out the the session we begin whit the login and the password and to avoid that another user can accede with our information, is very important that when we stop adding publications or we have to leave our task, we must select the red link, this will help to the general security of the system and thus to guarantee that the information in is reliable. In this page it is asked by if you want to add publications from the automatic index either in manual or automatic way. If we select the automatic option all the publications from the automatic index will go directly to the temporary database where there are all those publications that were not check. When that process has, finished a window telling us that the process was successfully completed will appear. Although by default the addition process is manual, this option will be explained next.

10. After selecting an option press the “Yes, Add Publications from the automatic index” button.
11. Then the publication will be appearing one by one, in order that the user can be able to correct those fields where the information is either incorrect or incomplete, this process is similar to the “add publications” exposed before. After checking every thing we shall press the button for adding to the database and all the following process is identical to that of adding a publication.
12. After adding each publication, its references, if it has them, and the author of the same one, are added.
13. At the end, when there are not more publications to be added, the user is notified about it and invited to leave the session.
14. This finishes the “Add from the automatic index”, to the database, process.

Chapter 12

Understanding the error messages

Different messages of error can appear in the use of ScienSe, in this section we want to explain to you the reasons that explain the why these errors happen, in order to avoid them, when possible, in the future. We present them next:

1. “Author Exist In The Database System”: It means that you are trying to add an author who is already in the database.
2. “This Citation already exist in the database system”: It means that you are trying to add to the database a reference that is already registered.
3. “Other user in the system or htsciense obtaining the articles from the web”: Another user with level of access as administrator, is within the system or htsciense is indexing publications. If you try to accede to the system later, the message surely will be disappeared, in opposite case contact the administrator of the system or if you are the administrator, review the log files.
4. “The automatic index does not exist”: This means that htsciense has not initiated its work or that it has not added more publications from the last time that you made the revision. If this message persists, contact with the system administrator.
5. “The Key Of Author is Invalid”: This happens because in some operation that requires the key of the author you introduced an incorrect

value as a empty field or a letter instead of a number. Generally this message only appears if either your browser does not have javascript support or it is deactivated or you want to send an invalid form on purpose, if it is this the last case we invite you to do not to do it.

6. “The Key Of Publication Cited is Invalid”: This happens because in some operation that requires the key of the author you introduced an invalid value as an empty field or a letter instead of a number. Generally this message only appears when either your browser does not have javascript support, or it is deactivated, or you try to send an invalid form deliberately, if this is the last case we invite you to do not to do it.
7. “The Key Of Publication is Invalid”: This happens when in some operation that requires the key of the author you introduced an erroneous value as a empty field or a letter instead of a number. In general this message only appears when either your browser does not have javascript support, or it is deactivated, or you try to send an invalid form deliberately, if this is the last case we invite you to do not to do it.
8. “The Month Of Publication is Invalid”: This message is display only when you try to send an invalid form on purpose, we invite you to do not make these kind of actions.
9. “The name of author is Invalid”: This happens because in some operation that requires the name of the author you left the field empty. In general this message only appears when either your browser does not have javascript support, or it is deactivated, or you try to send an invalid form deliberately, if this is the last case we invite you to do not to do it.
10. “The Title Of Publication is Invalid”: This happens because in some operation that it requires the title of the publication you left the field empty. In general this message only appears when either your browser does not have javascript support, or it is deactivated, or you try to send an invalid form deliberately, if this is the last case we invite you to do not to do it.
11. “The URL Request is Invalid”: This case happens when you modify the URL of some “get” request or try to make a “get” request from a

page that does not support this method. We invite you to do not to do it.

12. “The Year Of Publication is Invalid”: This error appears because in any operation that requires the year you introduce a year smaller than zero or higher to the present year. In general this message only appears when either your browser does not have javascript support, or it is deactivated, or you try to send an invalid form deliberately, if this is the last case we invite you to do not to do it.
13. “The Key does not exist in the database system”: It appears when in an operation that requires the previous existence of a key you give one that does not exist in the database system. Although it can also appear if you try to send an invalid form deliberately, if this is the case we invite you to do not to do it.
14. “Available Page or resource is not in this moment”: This type of messages appear when the server has a great amount of users and it cannot take care of your request at this moment. If this type of message appears very frequently we invited you to contact with the system administrator.
15. “Page or resource not found in this server”: It appears when you ask for a nonexistent resource in the server, we invited you to do not ask for pages that cannot be acceded from the main index.
16. “This Publication already exist in the database system”: It happens when you try to add a publication that already exist in the database.
17. “Text to Find is Missing”: This kind of error appears when you want to make a search and you left the text field empty. In general this message only appears when either your browser does not have javascript support, or it is deactivated, or you try to send an invalid form deliberately, if this is the last case we invite you to do not to do it.

Chapter 13

Modifying the information of the database

At this moment there is no support for modifications to the database through the web interface. Only the system administrator can make it directly by using either a SQL operation in the console of the database system or the SQL Tester software. If you need to make a modification to the data contained in the data base, please to contact yourself with the system administrator, and if you are administrator, please read the section titled “Guide of the Administrator”.

Chapter 14

Getting help and support

There exist different ways in which you can obtain aid and support for Sciense, we listed them in the order in which we considered that you should use them to facilitate the work for finding the answer.

1. Search in the contextual help: In generally each part of the application counts with contextual help, and considering that we have worked so that the application can be of simplest use as possible, there are great possibilities that your doubt can be solved checking the contextual help.
2. If you could not find the answer to your doubt, you can review the FAQ, where it is possible that you find a solution to your problem or doubt.
3. In case that you have not found the solution yet, you can review the help section that comes with the system so as to see if you find there the answer to your question.
4. The following step is to contact yourself with the administrator of your system in order to see if he can help you.
5. If either the previous thing is not possible or you are the administrator or your problem could not be solved then visit the web page of the project <http://sciense.sourceforge.net> and go to the help and support section, probably you will find the solution to your problem there.
6. If in that point you did not find solution you should ask in the forums or mail list of support so as to see if somebody else can help you.

7. If all this fail do not get hopeless, try to solve the problem by yourself, consider that with ScienSe you have access to the source code such as you are able to solve the problems that appear. And report the problem or error to the list of errors in order to be able to give solution as quick as possible, this way you will be cooperating with the improvement of this software.

Chapter 15

Contributing with the project

There exist many ways in which a user can contribute with this project. Between those they are:

1. Reporting errors found in the software.
2. Reporting errors found in the help.
3. Giving suggestions for improving ScienSe.
4. Participating in the support forums.
5. Implementing improvements in ScienSe.
6. Writing documentation.
7. Programming plugins.
8. Writing themes for the web interface.
9. Helping other users of ScienSe.
10. Writing packages for different platforms.

For a greater reference go to the page of the project in the developer section.

Part III
Administrator's Guide

Thanks for using ScienSe (SCIENtific SEarcher), we hoped that it will be useful and a help in your research work. We also, want to invite you to contribute with the development of this project either directly with code, giving new ideas for the software, reporting errors or saying to us what you like and what you dislike of ScienSe, with these you will contribute to make of ScienSe a better system.

Chapter 16

System Requirements

When ScienSe was developed, we had in mind that it must work by using the minimum of requirements of hardware and software, and also that the software requirements could be obtained as easily as the ScienSe itself. We also wanted that every one of the software requirements had as free version as a proprietary option in order to have several options for being able to choose, although for simplifying the tests we chose some defaults.

In this guide we will show you how to install, configure and maintain the system in optimal operation. Each one of the steps is oriented to make the things as simple as possible, nevertheless given to the great amount of possible systems and configurations any problem might exist, if this were the case we invited you to review the FAQ, to ask for support by an electronic mail in the support forums or if you, yourself solved the problem to write telling us how you did it.

This guide is done thinking about systems with software by default and considering as operating system a Linux with RPM support.

This is motivated by time causes, since it was in platforms of this type in which we tested the software. For future versions of this guide we will add aid for other configurations, in particular we are interested in packages for systems based on Slackware, Debian and FreeBSD.

Before installing ScienSe it is necessary to ask us about what we want to do with ScienSe, because the requirements of the system will vary if you

want to install ScienSe in a server, to participate in the development, to review the design, to adapt the system or to build ScienSe from the source code.

Once you have identified what you want to do with ScienSe, you should review if your system has the software and hardware requirements or in other case, make the necessary changes for fulfill the requirements.

16.1 Hardware

The hardware requirements depend so much on the following parameters:

1. The selected components of software.
2. The amount of sites to be indexed.
3. The amount of possible users.
4. The amount of data for the stored publications.

And as the system has been designed thinking about the portability, ScienSe can work practically in any platform that can execute a J2EE implementation and counts with POSIX support.

16.2 Software

Next we enumerate the necessary software for the different things that you might want to do with ScienSe. It is possible to indicate that first there are the default options in which ScienSe has been proved and after we give the alternatives within the commercial and free/open software.

16.2.1 For executing ScienSe

- **Java Development Kit 1,4 or a newer version:** It can be used any Java compatible development environment.
- **Apache Web Server 1,3 or a newer version:** It can be used any web server with CGI support in C.

- **Apache Tomcat 5,0 or a newer version:** It can be used any compatible server with Java 2 Enterprise Edition.
- **Linux Operating System kernel 2,4 or a newer version:** It can be used any system compatible with POSIX and that fulfills the other requirements of software, as: FreeBSD, OpenBSD, GNU Darwin, MacOS X, Solaris, Windows 2003, etc.
- **PostgreSQL 7.1.2 or a newer version:** It can be used any database manager system, SQL92 compatible as Firebird, MySQL, Oracle, DB2, etc.
- **A HTML viewer with frames support:** As Netscape, Mozilla, Konqueror, Operate, Galeon, Epiphany, Firefox, Safari, Internet Explorer, etc. We suggest you a browser with javascript and CSS support, although they are not essential.

16.2.2 In order to see the documentation

- **A HTML viewer.**
- **A PDF viewer:** Like XPDF, Acrobat Reader, Ghost View, etc.
- **Optionally OpenOffice 1,0 or a more recent version:** StarOffice 7,0 or superior can be used also.

16.2.3 For compiling ScienSe

- **Java 2 Software Development Kit version 1,4 or superior:** It can be used any Java compatible development environment.
- **GNU Make.**
- **Ant 1,5 or higher.**
- **GNU C++ Compiler 2.95 or a newer version:** Any ANSI compatible compiler can be used.
- **autoconf.**
- **automake.**

- **J2EE library:** Those of Tomcat can be used.
- **Optionally if you desire to make RPM packages, the tools to create them.**

16.2.4 For modifying the documentation

1. **Open Office 1.0 or a newer version:** Also StarOffice 7,0 or superior can be used.
2. **Dia 0,88 or a newer version.**
3. **ÆT_EXSoftware.**
4. **Argon UML.**

16.2.5 For participating in the ScienSe project

- **The same requirements than for compiling and modifying the documentation.**
- **The control versions system CVS.**

Chapter 17

Preparing the environment

After analyzing the software and hardware requirements, the following step is to obtain the necessary software. Most of the requirements are already installed in your system if you chose to use the default platform. In this section we will only talk about the components that in our opinion are the most difficult to obtain or that are not available in most of the systems and again taking as bases the default components.

If you wish to install other components and find some problem, we recommended you to visit our Web site in the help and support area, it is possible that there you can find help about that matter. We would like also that in case that you install successfully, the system in platforms where we have not made tests, to notify us in order to add your experiences in the help and support section.

- J2SE Software Development Kit (SDK) 1.4.2 or superior

It can be obtained from the following address: <http://java.sun.com/j2se>

- Apache web server 1,3 or higher

It can be obtained from <http://httpd.apache.org/> although if you use a Linux or FreeBSD system we recommend you to use the package that comes in your distribution.

- Apache Tomcat 5,0 or superior

It can be obtained from <http://jakarta.apache.org/tomcat/> or from <http://www.jpackage.org/> if you use RPMs. If you already have installed it in your system we recommend you to use it and in other case

we recommend you, if your system supports RPM, to use the RPM package that we provided along with ScienSe.

- Ant 1,5 or superior

It be obtained from <http://ant.apache.org/> or from <http://www.jpackage.org/> if you use RPM. If it is already installed in your system we recommend you to use it, and in opposite case if your system supports RPM we recommended you to use the `install_ant.sh` script that is in the project's web page, which is in charge of obtaining it through Internet and installing ant and all its dependencies.

- PostgreSQL 7,3 or higher

It can be obtained from <http://www.postgresql.org/>. But if you already has it installed in your system we recommend you to use it.

17.1 Installing the software required

The procedure of installation of the ScienSe requirements varied depending on your system and the package, by this reason in the case of the requirements we recommend you to read the help of each package in particular.

In the case of ScienSe, this is distributed in form of a RPM package, although you can also obtain the source code in a compressed file or through CVS. In the future we will have available more types of packages.

17.2 Configuring PostgreSQL

In the case of PostgreSQL there are 2 important things to configure; the first is to create the database for ScienSe, and the second is to allow connections JDBC from ScienSe to PostgreSQL.

Setting Up PostgreSQL:

1. If you already have databases working in your system that allow JDBC connections go to the next point. In other case continue to the next interjection.

2. adduser postgres.
3. mkdir /var/lib/pgsql/data
4. chown postgres /var/lib/pgsql/data
5. su - postgres
6. initdb -D /var/lib/pgsql/data
7. postmaster -D /var/lib/pgsql/data > logfile 2 > & 1 &
8. createdb test

Allowing JDBC connections

1. In the file postgresql.conf to add the line: tcpip_socket = true
2. In the file pg_hba.conf to add the line:

```
host all all ip_server net_mask trust
```

where ip_server is the IP direction of the server where is tomcat and net_mask is the net mask.

17.3 Configuring Apache

Generally ScienSe works well with the standard configurations of the Apache web server. The most important option for the operation of ScienSe is that apache should allow the execution of CGI scripts. Thus file httpd.conf must have the following line:

```
LoadModule cgi_module modules/mod_cgi.so
```

17.4 Configuring htsciense

Htsciense is a software based on htdig (<http://www.htdig.org>) that carries out two fundamental functions:

1. It creates an index with the indicated sites and stores that information in a Berkeley database in order to make searches by means of a cgi script, later.
2. It looks for research articles and analyzes them so as to extract certain information, after that, it stores those data in a relational database.

As can be seen, the most important information that is needed by htscience to work, is the sites to be reviewed, although many other configurations can be supported. At the moment of writing this document, it was being developed htConf, an Java Swing application that facilitates the configuration of htscience by means of a friendly graphical user interface.

The htscience configuration is carried out writing a configuration file htscience.conf that is located in the directory “/etc” on UNIX systems. The installation of ScienSe already contains a default file htscience.conf with a configuration that will have to be used for most of the cases, the only necessary thing is to change the option “start_url” where shall be the sites to be reviewed.

The rundig instruction can be executed, so as to begin the indexed process, although it is not necessary, because in the installation process, a new entry is added to the cron command, in order to make automatically the process, every day at middle night. That value of cron, can be modified if you desired it.

17.5 Configuring Tomcat

ScienSe works practically with the default configuration of Tomcat, the only thing to be modified is the role “science” and a user with that role in the configuration file; tomcat-users.xml, this user will be the one that will have the possibility of reviewing and of adding publications automatically obtained.

In order to add the roll “science” in the file tomcat-users.xml (that generally is in /etc/tomcat5) you should add a line as the following:

```
<role rolename=”science” />
```

And for adding a user with this roll in the same file a line like the following one should be added:

```
<user username="name" password="pass" roles="science"/>
```


Chapter 18

Installing ScienSe

The installation depends on if you got ScienSe in a war or a rpm format.

If the application was gotten in a rpm format, you should executes the following command:

```
rpm - ihv htsciense-1.0-i386.rpm sciense-1.0-noarch.rpm
```

and the application will be installed and working, if you want to check it, you only have to open the application in tomcat with the context path /sciense.

If the application was obtained in a war format you must follow these steps.

1. Open the Tomcat manager in the following url:

```
http://localhost:8008/manager/html
```

Note: Port 8008 can vary depending on the Tomcat installation.

2. Go to the option of “War file to display”, after that select the sciense.war file and press the button “display”.
3. The application will be installed in the context trajectory /sciense.

In general, the default configuration values of the ScienSe application will work well in most of the cases, although they can be modified so as to adapt them to your requirements. Most of the configurations related to Tomcat are located in the files “/science/web/WEB-INF/web.xml” and “/science/web/META-INF/context.xml”.

The parameter that must be changed is the related one with the access password to the database that is in the file context.xml.

Chapter 19

Processes executed by ScienSe

It is important that the system administrator knows about the processes executed by this software, so that in the case that one suspects a intrusion in the operating system, he can detect which are not the normal processes executed by this program, moreover being an application as ScienSe which complexity requires of the execution, sometimes in simultaneous form, of a great amount of programs. The processes that are executed with ScienSe are the following ones:

1. **postmaster.** The postgresql server.
2. **httpd.** The Web server apache.
3. **Java:** The tomcat server.
4. **wget:** Is encharge to download articles of the network to be analyzed.
5. **htsearch:** A CGI program for searching in the index.
6. **Htload, htsciense, htdump, htmerge, rundig, htfuzzy, htnotify:** Applications of the htsciense suite.

Chapter 20

The indexing process

One of the most important steps in order to be able to maintain the system working in a good state is to understand each one of the steps that the program executes to make its work, between these the indexed process is one of the most important. The indexed process works in the following way:

In the cron demon is registered the rundig script, which executes every week at night in order to make the following process. The rundig script executes several processes that are in charge of creating the index of the sites and to analyze the publications that are there, between these processes htsciense is the most important. htsciense enters to the sites given by the user in the htsciense.conf file and begins the indexed process. This it consists of creating an index with the words found in the document, following recursively the links that it found, to analyze the information of the document for getting the information of the publications and to store this information in a text file that will be used for filling the database later.

When htsciense finds a document in a HTML, PDF, PS, DOC, etc, format, calls an external process that executes a filter for transforming the document into a flat text and places the data in a temporary file, from which soon the data are read so as to be analyzed and extract their most important information.

The algorithm used, for analyzing these files, is based on heuristics that use probabilities.

While htsciense is making the indexed, the data file where it places the information of the identified scientific publications is blocked, with a cooperative blocked so as to avoid conditions of competition with other ScienSe's processes.

The information of these publications is stored in a format of tags that can be read by a web application.

Chapter 21

The checking process

Once htsciense has carried out the indexed process, it begins to store the data collected to the relational database. This process can be made with or without revision. The revision must be carried out, because the algorithm that identifies the information of the publications could not find the correct data always, due to the articles in the network are in a great variety of formats.

When the revision takes place, the publications are stored in the database for their later consultation, although it also exists the option to store the data without revision, this can be required, for example when we do not have enough time for making that revision and we want to obtain benefit of the information as rapidly as possible. It is possible to indicate that when the data have not been reviewed, these are stored in a temporary database, previous revision, can go to the database with the reviewed information.

With this we get several advantages of the automatic indexed, but at the same time it is guaranteed that the information of the database is reliable.

Chapter 22

The adding information process

Aside from the automatically data obtained, more information can be introduced by means of a web format. This information is stored in the database of reviewed publications. This application is connected by JDBC to the database so as to store the information.

Chapter 23

The searching process

The search of information is carried out by means of forms of a web application. This application is connected by JDBC to the database in order to obtain the required data.

Chapter 24

Maintenance

The maintenance to the system consists mainly of carrying out the process of revision previously described and of backing up the information, process that we will explained in the following section.

Chapter 25

Backing up the information

The backing up of information consists of backing up the database and the configuration files. The configuration files of ScienSe are web.xml, context.xml and htsciense.conf. In order to backing up the database please read the manuals of the database system that you are using, generally this will be PostgreSQL.

Chapter 26

Getting support

You can get support for ScienSe in several ways, we recommend you to follow the next steps, if you find any problem.

1. To read this guide, it is probable that here you might find the solution to your problem.
2. To review the FAQ in the administration section.
3. To visit the projects' web page, <http://sciense.sourceforge.net>, and to go to the help and support section, probably, you will find the solution to your problems there.
4. If in that point you did not find the solution, then ask in the forums or support mail lists so as to see if somebody else can help you.
5. If it is not, do not get hopeless, try to solve the problem by yourself, consider that with ScienSe you have the access to the source code, thus you are able to solve the problems that appear. And report the problem or error to the errors list in order to be able of giving solution as quickly as possible, in this way you will be cooperating with the improvement of this software.

Part IV
Programmer's Guide

Chapter 27

Basic structure

ScienSe is a system developed under the client-server model. In the client's side, it has a web interface, based on HTML, CSS and Javascript, that allows to manage the information contained in the system. HTML is used to organize the information of menus, forms and presentation tables. CSS is in charge of the graphical presentation to the user and with Javascript it validates the data of input and presents some elements which makes the interaction between the user with the application, easier.

On the server's side, the system has 3 main components:

The first of them is a relational database that stores information of scientific publications, the database, and all the statements that manipulate the elements contained in the database were developed using SQL92 so as to get the independence of the database manager system.

The second component is an web application in the server's side which allows to manage the information contained in the database. This application was developed using Java as programming language and the J2EE, Java Servlets and Java Server Pages technologies. The connection with the database is made by through of JDBC.

The third component is a program that creates an index of the web sites given by the user, looks for scientific publication information and stores these data in order to add them to the database later, after a previous revision of the user. It was developed using ANSI/ISO C/C++ and it makes use of the

interface libraries with the POSIX system.

Chapter 28

APIs

ScienSe counts on a documented API with the javadoc system which allows to make extensions to the system so as to adapt it to the users' requirements. In a future it will also have APIs for the configuration, the plug-ins development and the extension of htsciense.

Chapter 29

Converters

Converters can be developed that turn files from different formats to ascii text that can be analyzed in the simplest way so as to be able of getting the information about the publications.

Chapter 30

Plug-ins

At the present time, ScienSe does not have plug-ins support but in future versions it will count with plug-ins support so that the programmers can develop and integrate improvements on the algorithm of getting information of publications, in a easy an transparent way.

Chapter 31

Web Interface

The web interface uses HTML, CSS and Javascript for the validation. At the present time the interface is in the servlets, thus, in order to be able to modify this interface, it is necessary to know the Java programming language and in some cases the operation the Java Servlets also, in a future we will try to separate the presentation from the logic being allowed to modify the web interface with no need of knowing these technologies.

Another important point is that the web interface must be designed with base on open standards which will allow it to be visible in almost all of the web browsers available. So far we used CSS 1,0, HTML 4.0 and Javascript 1.5.

Chapter 32

The database

It is recommended not to modify the database so as to avoid to break its integrity and to offer compatibility with the standard application and thus be able of getting support in a simpler way. Within the programmer's documentation there are the tables and the diagrams entity-relation that will allow the programmer who wishes to extend ScienSe be able of designing correct SQL sentences.

It is necessary to emphasize that we counted with JDBC Tester, a Java Swing application that let you prove JDBC connections and SQL sentences. This tool will make easier the work of extending ScienSe or any other application based on JDBC, to the programmer.

Chapter 33

Security

As ScienSe is an application for networking, the security is one of the things that we have taken more in consideration during the design and the development process. For that reason we have pay special emphasis to the validation of data, the permissions and the safe programming.

Part V
Developer's Guide

This guide is thought so as to orient to that programmer who is interested in integrating to the development team or simply to anyone who wants to know more about the details of Sciense.

Chapter 34

Details of design

ScienSe is a system developed under the client-server model. On the client's side there is a web interface based on HTML, CSS and Javascript which allows to manage the information stored in the system. HTML is used in order to organize the information of the menus, forms and presentation tables. CSS is in charge of the graphical presentation to the user and Javascript is used to validate the input data and for present certain elements make easier the interaction between the user and the application. This validation of the data is made only for improving the user interface, since by security reasons within the components that manipulate the information those validations are made again. All the interface is built-in on the servlets, thus, in order to modify the interface, it is necessary to know this technology.

On the server's side the system consists of 3 main components:

The first of them is a relational database that stores information about the research publications, the database, and all the statements that manipulate the elements contained in the database were developed using SQL92 so as to allow an independence from the database manager system. Within the source code of the application are the tables and the entity-relation diagrams of the database.

The second component is a Web application on the server's side which allows to manipulate the information contained on the database. This application was developed using Java program language and technologies as J2EE, Java Servlets and Java Server Pages. The connection with the database is

made by means of JDBC.

The third component is a program that generates an index of the Web sites given by the user, there it looks for the information of scientific publication and stores that information so as to add it to the database, later, after being checked by the user. If it finds files in formats as HTML, PDF, PS, DOC, among others, it uses a filter in order to turn them into ASCII text and then apply the identification of data algorithm. All the generated information is stored in a temporary file, which when the indexation process has finished, is read by the web application in the check publication process. wget is used for download the documents in order to analyze them. It was developed using ANSI/ISO C/C++ and uses the system POSIX's interface libraries.

Chapter 35

Development tools

In the development of ScienSe we have only used free development tools, with the purpose of making accessible to the programmers to participate in this project. Between the tools that we used are autoconf, automake, GNU make, ant, javadoc, Doc++, gdb, CVS, BlueFish, etc. And the languages Java, C/C++, Javascript, HTML, CSS and SQL92.

Chapter 36

HTSciense

This program is based on htdig version 3,6, in this version has solely been made modifications in the user interface and some adaptations to allow the process of analyze publications. In a future there will be made some improvements so as to optimize its work.

Chapter 37

The database

The database is relational. It has been standardized and has mechanisms to verify the referential integrity. For its creation there was used only statements of the standar SQL92, in order to be able to ensure the portability of the system.

Chapter 38

The web application

The web application is developed following the design patterns for the J2EE applications. It uses JavaServlets and Java Server Pages for the automatic generation of content, Javascript as for the validation of the data as to make richer the user interface and finally, CSS is used for improving the interface.

Also it is used a tail of connections and handling of concurrence in the database accesses. In order to make these connections JDBC is used.

Chapter 39

Security details

Aside from using the security mechanisms of the objective operating system and from the applications that Sciense uses, within the application we have thought about the security of the system.

Being the problems of buffer overload, the memory leak and the send of invalid data three of the main causes that harm the security of the systems, we have paid special emphasis on these points.

With respect to the buffer overload and the memory leak we used Java in the web application, this language has mechanisms of Exceptions and Handling of memory that diminish these problems. As well, we have meticulously reviewed the memory leak in htsciense that is written in C++ and we used the class string of the Standard Template Library, which also contains mechanisms that help with the buffers overload and memory problems.

With respect to the send of invalid data, in the web application, we validated the data with Javascript for making the web interface richer and also we validated the data at the application level, anticipating with that the send of invalid data by modifications to the forms, or re-writings of urls. In htsciense, the input data are verified again and there are mechanisms that allow to control the modifications or errors in its execution environment.

Chapter 40

The documentation

One of the policies of ScienSe's development is to document all the source code so as to facilitate the revision and modification to this software. The documentation process should be made following the Javadoc standard of documentation for the code as in Java as in C++, since the use of this tool allows the automatic generation of documentation in-line for the diverse APIs.

In addition to the documentation of the source code, it is important to count with all the documentation about the development process using the UML language. Also the design of the database should be documented, by the tables and the entity-relation diagrams so as to make easier the queries design process.

The user documentation should be divided in two parts, on one hand, the documentation that contains the application should not contain information too much specific that can tend to change with facility. This type of information must be added to the help and support section of the ScienSe's Web site.

Chapter 41

Future Projects

Of course the ScienSe's development team is conscious about the weak point of the product, thus thinking about them, we present the following list of characteristics that could help to eradicate of such weaknesses.

1. The first thing considered of vital importance is reinforcing the extracting information algorithm, since the one that is implemented in this version, present some fails in front of certain articles with outside the ordinary format, because it is based on probabilities. We have thought that an alternative to this algorithm is to use a trained neural network, since they have demonstrated in numerous occasions ideal characteristic for this kind of work.
2. Another improvement that could hit in a significant way the usability of the system is the creation of a friendly graphical user interface for the configuration of the indexer since, in this moment, this process is through a configuration file that, in spite of being of easy creation and modification, forces the end user to submerge in a sea of information that would be not so useful for him, even more when he requires to make things as easy as add a new url to the searching. The creation of an interface that contained only the variables that the user should configure and which can be able to add those values to the indexer's configuration file, would allow an even more easier interaction of the user with the program. At the present time we have a version of such application, called htconf but it still in a development phase. It would also be desirable to allow the configuration of the system through the Web.

3. Another one of the things that are being developed is JDBC Tester, a tool to prove JDBC connections and SQL statements, by means of a friendly user interface. At the present time it is already functional but it still need some improvements in the presentation of the results and in the system's help.
4. Adding the possibility of doing combined searches is another one of the improvements that are consider, for this it would be necessary to modify the components that access to the database and to add code to the interface, so that facilitate to the user's the searching process.
5. When htsciense finds files with the formats: PDF, HTML, DOC, PS, etc., creates a process that calls a filter that makes the conversion of the given format into plain text. These filters need to be improved and to developed more so as to handle other formats.
6. In the future we will find that for different types from publications certain extracting information algorithms are better than others, in addition of there will be who want to prove a new algorithm without need to modify all the source code of htsciense. Therefore we believe that is important to add the plug-ins support in order to allow the use of different algorithms with neither any needed of compiled the application each time nor knowing the details of its operation.
7. In order to improve the user's participation, we think that is important give them the possibility of modifying the information of their own publications at any moment, as well as to add new data. In such a way that each author can participate actively in the getting information process so as to make the system more valuable. Therefore, adding user's sessions would allow to improve as the interface as the trustworthiness of ScienSe.
8. In order to allow that ScienSe can adapt better to the requirements of different users without having to invest long time and resources in adapting the system to the particular necessities, we want to introduce in ScienSe the Model-View/Controller in the application, in order to separate the logic from the presentation. In this way the modification of the interface won't require of programming knowledge.

9. The use of frames was for improving the user interface. But later we realized that it could have problems in those browsers without frames support, therefore we believe advisable to modify the application so that it does not use frames.
10. There exist some data that can be useful to know about a scientific publication, but even though that information can already be obtained from the system, there is no a place within the interface that shows this results to the user. Between the data that could be of interest there are: the most referenced authors, the most referenced publications. Thus a possible improvement, would be to add a section in the web interface which shows this information.
11. Once it has the sessions' support, it can be able that the users could choose between different web interfaces, depending on which is the one that better adapts to his necessities, this can be done by means of themes support in the interface.
12. Given the few resources with which we made the project, there was not possible to test the system in different platforms, therefore it would be desirable to prove the program in different platforms of operating system like MacOS X, FreeBSD, Solaris, Debian, Slackware, for mentioning some. Besides to prove the system in different implementations of J2EE, in other database manager systems as MySQL, Oracle, Firebird. And to elaborate packages for the different platforms.
13. Another one of the improvements that can make of ScienSe a better tool for working and could make easier to get its maximum benefit would be to add support for JavaHelp, so as to have a powerful and consistent help system.
14. We wish to improve the application logo also, thus we hoped that some user with knowledge of graphical design can help us. The unique requisite is to make it using The Gimp.

Part VI
References

Chapter 42

Java Servlets

Servlets or Java Servlets is a technology developed by Sun Microsystems company (<http://www.sun.com>) which provide a method based on components and platform independent for the development of applications based on the Web, without the limitations of cgi programs. And without the extension mechanisms of proprietary server, servlets are also independent of the server. This gives the freedom for choosing server, platforms and development tools.

Chapter 43

Java Server Pages

The Java Server Pages (JSP) are an extension of the Java Servlets, this technology, also developed by Sun Microsystems, allows to the programmers and web designers, to develop and to maintain quickly, dynamic pages rich in information. As part of the family of Java technologies, the Java Server Pages allow the development of applications based on the Web that are platform independent. JSP technology also, separates the user interface of the generation of content, allowing to the designers to change the page design without altering the underlying dynamic content.

Chapter 44

Javascript

Javascript is an object oriented scripts language of crossed platform of Netscape. Javascript is a small and light language; it is not useful as an independent language, but it is designed for being easily embedded in other products and applications, such as web browsers.

It is used generally for providing richer user interfaces and for validating the input in forms of web applications.

Chapter 45

Tomcat

Tomcat is a container of servlet that is used as the official implementation of reference for the Java Servlet and Java Server Pages technologies. Tomcat is an open software distributed under the license “Apache Software License”.

Chapter 46

Apache

Apache is the most popular web server in Internet with almost 70% of the market. It implements the most recent web protocols, is highly configurable and tensile by means of an API for the development of modules, runs in the amplest range of platforms, besides to allow the access to all its source code.

Chapter 47

Postgresql

PostgreSQL is a relational database manager system based on postgres version 4.2, developed in the department of Computer Sciences in the University of California at Berkeley. Postgres was a pioneer in many concepts that appeared much later in some of the commercial database systems.

PostgreSQL is an open source descendant of the original code of Berkeley. It supports SQL92 and SQL99 in addition to other modern characteristics, among them:

- Complex queries
- Foreign keys.
- Triggers
- Views
- Transactional integrity.
- Concurrency control multi-version.

Also, PostgreSQL can be extended by the user in several forms, for example adding:

- Data types
- Functions

- Operators
- Aggregation functions.
- Procedure languages.

And because of its license, PostgreSQL can be used, modified and distributed by anyone free of position for any intention, private, commercial or academic.

Chapter 48

GNU Wget

GNU wget is a free software package for getting files using the most used web protocols as: HTTP, HTTPS and FTP. It is a non-interactive tool of command line, which can be called from scripts, cron tasks, terminals without X support, etc.

wget has manifold characteristic to easily get large files or entire Web sites or FTP sites copies.

Chapter 49

RPM

The RedHat Package Manager (RPM) is a powerful tool of command line developed by the RedHat company (<http://www.redhat.com>), packages control manager system, able to install, uninstall, verify, consult and update software packages. Each software package consists of a set of files in a package with information about the package like its version, description, etc. There is also an API (Application Program Interface), that allows to the advanced programmers to pass these commands and to handle these transactions in a native language.

Part VII

FAQs

Chapter 50

General

1.1 What is ScienSe?

ScienSe is a scientific research tool of support. ScienSe looks for, organizes and shows the fundamental information of the research articles and other publications of scientific character.

1.2 What is the meaning of “ScienSe”?

ScienSe is an abbreviation of Scientific Searcher that is the long name of the system.

1.3 What is the source of the name?

The election of the name was one of the most important processes in the project due to the expectations that were had. The name had to express the essence of the project, in addition to being specially attractive to the public to whom it is directed. Long time went by before it was happened to us, but when it came us to the mind, we had the certainty that if there existed a name for this software, it had to be SCIENTific SEarcher.

1.4 What is the license of ScienSe?

Of all the range of attractive characteristics, the license is one of most important, because there is not any other software with the characteristics of ScienSe in the market that, like ScienSe, is under the general public license (GPL). This means that you will be able to have access to the source code, and even more, you will be able to modify, adapt or improve it, according to your requirements.

1.5 In which operating system does ScienSe run?

ScienSe has been developed and proven in the Linux platform, although in principle it would have to work without problems in any operating system with POSIX support and a J2EE implementation.

1.6 Where can I get ScienSe?

ScienSe is available in several packages in the download section of the following direction:

<http://science.sourceforge.net>

In addition if you wish to prove its potential before download it, we already have a web site running with ScienSe called ScienSeMath, that is on the following URL:

<http://itza.fciencias.unam.mx/sciensemh>

In spite of in ScienSeMath you will have access only as an extern user, it will give you a great vision about the things that ScienSe offers you. ScienSeMath contains the information of the mathematics research articles of the most important magazines on the area.

1.7 Where do I get support?

We have carefully documented each aspect related to ScienSe, thus there are several options.

- (a) The contextual help that is integrated on the application.
- (b) This FAQ.
- (c) The user's, administrator's, programmer's and developer's manuals.
- (d) In the help and support section of the ScienSe's page.
- (e) In the support lists.

1.8 What is the documentation available?

The user's, administrator's, programmer's and developer's manuals, the FAQ, a help and support section in the ScienSe's web site, the API documentation and the documentation of the development process that is in the source code.

1.9 How could I Know about the bugs or the missed characteristics?

The better place for this, is the help and support section of the ScienSe's project page.

1.10 How can I access to the ScienSe's development team?

In the developer's area of the ScienSe's site there are information about the matter.

1.11 How can I report an error?

In the help and support section of the project's site there are indications about how to report a bug.

1.12 How is ScienSe compared with the other software of the area?

It has almost all the characteristics of other systems with the advantage that ScienSe is under the GPL license and free.

1.13 How can I help to finance ScienSe?

There is information on the matter in the "About ScienSe" section of the main page of the project.

Chapter 51

User

- 2.1 Which are the features that my browser should have in order to enjoy all the interface qualities?

Although you can access to the system with any browser with frames support, the system will work better if your browser supports Javascript and CSS.

- 2.2 Would I add my own information by myself?

Yes you would, in the user's manual you can get the necessary information for adding authors publications and references.

- 2.3 Would I get an interface in my native language?

No and there are no plans so that this happens, mainly because almost all the scientific information appears in English language therefore we think that it is not necessary to offer the interface in other languages.

- 2.4 Why ScienSe does not identify all the information of the articles correctly?

There exits many reason by which ScienSe is not able of identify all the information of an article correctly, the most commons are:

- (a) The format of article is not common. ScienSe tries to identify all the information of an article using probabilities, so, it is designing for identify the information in the most common format, when this is not true, ScienSe can not know which is the sort of the fields, so it gets confused. Although this is not a problem that affects the

most quantity of the articles we hope to improve notably by using a neural network (see Future Projects for more information about it).

- (b) The programs that helps to convert the file, from any of the supported formats, to text can no convert the file of the article correctly. ScienSe can support files with different formats as pdf, ps, rtf and text, but in order to work with a file, ScienSe requieres it to be as a plain text file, so it uses some format converters that get those file as plain text files. Unfortunately, there are some programs that save the files with some no standar characters or format, so the converters can not get all the information from the files with that characteristic. Thus ScienSe receives a damaged file different from the original.

2.5 I have install a new version of the converter filter, why does ScienSe seem to have changed? ScienSe can support different formats but in order to execute the algorithm for identify the information it uses some converters format (as pdftotext) so as to get all the files as plain text files. When you install a new version or a different converter, the plain text files might changed also, since the program which converts them have changed, so, the file given to ScienSe could be diferent, and this could change the results.

2.6 Which is the "ideal" format, that ScienSe requieres for getting the general information of an article? As there are a lot of different ways for writing an article, there is no way to be sure about the sort of the information, so in order to improve the performance of ScienSe is desirable that the files of the articles fulfill certain formats. The desirable format is:

- (a) The first line of the file should be the title of article, and it should finish with an endline (enter).
- (b) The authors should be separated by “,” (commas) except for the last two authors which should be separated by either the word “and” or the character & . For example: Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and Viso Elisa for meaning: 5 authors.

- (c) If the article contains the month of publication the entire word in english should be written instead of any abbreviation. For example: January, and no jan, or enero, or something like that.
- (d) The emails, if there is some, should be set in the same order that the authors where writen, and if there are some authores with no email address they should be set at the end. The email addresses can be separated by ',' (commas) or only with ' ' (spaces). For instance:

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and Viso Elisa fraggi@matem.unam.mx, urrutia@matem.unam.mx pvm@fciencias.unam.mx

for meaning:

Raggi Francisco fraggi@matem.unam.mx, Urrutia Jorge urrutia@matem.unam.mx, Miramontes P. pvm@fciencias.unam.mx, Bracho J. Viso Elisa

By simplicity there should be at most one email address by author.

- (e) The personal web page should be in the same order that the authors were given, and can be separater either by and ',' or a ' ' character. For instance:

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and Viso Elisa fraggi@matem.unam.mx, urrutia@matem.unam.mx pvm@fciencias.unam.mx
<http://www.matem.unam.mx/investigadores/raggi/> <http://www.matem.unam.mx/urrutia/> <http://www.mathmoo.unam.mx/biomat/pedro/pedro.html>,
<http://www.matem.unam.mx/investigadores/roli/>,

- (f) The abstract should begin with the word "abstract" and should finish with an end line character (enter). For instance:
 "Abstract. This is an example."
 or
 "abstract
 This is an example"
- (g) The phones should be a string of numbers with at least 5 numbers, separated by the character '-'. And they should be given in the same order that the authors. For instance:

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and
 Viso Elisa 50-00-00-00 51-11-11-11 01-23-45-56-78-9 98-76-54

for meaning:

Raggi Francisco 50-00-00-00,
 Urrutia Jorge 51-11-11-11,
 Miramontes P. 01-23-45-56-78-9,
 Bracho J. 98-76-54,
 Viso Elisa ,

- (h) The name of the journal or magazine should be given using any of the following words, and should finish with an end line (enter):
- i. "journal". Example: Linux Journal
 - ii. "magazine". Example: Science Magazine
 - iii. "communications". Example: ACM Comunications
 - iv. "scientific". Example: Scientific American
 - v. "proceeding". Example: Proceedings of the National Academy of Sciences
 - vi. "book". Example: Algebra's book
 - vii. "review". Example: ACM Operating Systems Review
 - viii. "nature". Example: nature.com
 - ix. "science". Example: Science Online
- (i) References section should be the last section of the article, and should begin with the word "references" (it does not matter if the letters are upper or lower).
- (j) Each reference should begin with one of the following ways:
- 1.- [anything] anything
 - 2.- (anything) anything
 - 3.- anything anything
 - 4.- anything.- anything
- Where anything means any character, number or letter different from the ones that contains it. NOTE: the first reference shows the way in which the rest of the references are going to be written, so if the first begins with [1] the next references should be given

with [and]. For example:

[1] Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p. [2] Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press
or

(1) Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p.

(MM) Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press, 34-40.

or

1 Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p.

a Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press, 34-40.

or

SM.- Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p.

1.- Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press, 34-40

2.7 .- Which is the desirable format for give the references of an article according with ScienSe?

Each cite should be given with the following format:

- (a) The pages of the journal where the article is, should be given either by two numbers separated by a '-' or by a number finished with "p." Example:

[SM] Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985, 300p. [P] Marvin L. Minsky and Seymour A. Papert, Perceptrons: Introduction to Computational Geometry, (1988), MIT Press, 34-40

- (b) The year of publication should be older to 1900 and younger to the present year (of the computer where ScienSe is running). The

year can be given as the number alone or as (number) with the '(' and ')' characters.

Example:

Marvin Minsky, "The Society of Mind", Simon and Shuter, 1985

for getting:

year=1985

or

Marvin L. Minsky and Seymour A. Papert,

Perceptrons: Introduction to Computational Geometry, (1988),

MIT Press

for getting:

year=1988

- (c) The journal should be the last string of the reference after erasing the year and the pages. The journal is consider as all the string before the last ',' (comma) character, so the journal should no contain any ',' character within the string.

Example:

Marvin L. Minsky and Seymour A. Papert,

Perceptrons: Introduction to Computational Geometry, MIT Press

for getting:

journal = "MIT Press"

- (d) The title should be the last string of the reference after erasing the year, pages and journal. The title is consider as all the string before the last ',' (comma) character, so the title should no contain any ',' character within the string.

Example:

Marvin L. Minsky and Seymour A. Papert,

Perceptrons: Introduction to Computational Geometry

for getting:

title=Perceptrons: Introduction to Computational Geometry

- (e) The authors are consider as contained in the string of the reference after delete from it the year, pages, journal and title. Each author is separated by the ',' character from the others and can

contain the word "and" or the character '&' separating the last two authors. NOTE: as it is suppose that each author is separated from the others by a ',' the any name of author shoul contain the ',' character. Example:

Marvin L. Minsky and Seymour A. Papert
or

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J. and
Viso Elisa

or

Raggi Francisco, Urrutia Jorge, Miramontes P., Bracho J., Viso
Elisa

2.8 Is the ideal format the unique format that ScienSe is able to work with?

No. ScienSe work reasonably good (it gets correctly up to the 50infor-
mation in an article) with almost any format, but the error's probabil-
ities increase when the articles do no have an usual format.

Bibliography

- [1] JavaServer Pages Specification; version 2.0; Proposed Final Draft 3. Copyright 2003 Sun Microsystems, Inc.; April 11, 2003.
- [2] Unified Modeling Language (UML) Specification: Infraestructure; version 2.0.
- [3] A Simplified Introduction to \LaTeX ; Harvey J. Greenberg; University of Colorado at Denver; April 18, 2000.
- [4] Maximum RPM; edition 2; Edward C. Bailey; Red Hat, Inc.; 2000.
- [5] Designing Enterprise Applications with the J2EE Platform; Second Edition; Inderjeet Singh, Beth Stearns, Mark Johnson, and the Enterprise Team; Addison-Wesley.
- [6] PostgreSQL 7.3.2 Administrators Guide; PostgreSQL Global development Group.
- [7] PostgreSQL 7.3.2 Programmers Guide; PostgreSQL Global development Group.
- [8] PostgreSQL 7.3.2 Users Guide; PostgreSQL Global development Group.
- [9] PostgreSQL 7.3.2 Reference Manual; PostgreSQL Global development Group.
- [10] Web Style Guide: Basic Design Principles for Creating Web Sites; 2nd Edition; Patrick J. Lynch, Sarah Horton; www.webstyleguide.com.
- [11] GNOME Human Interface Guidelines; The GNOME Usability Project; Calum Benson, adam Elman, Seth Nickell, Colin z Robertson.

- [12] Java Look And Feel Desing Guidelines; Second Edition; Sun Microsystems, Inc.
- [13] KDE User Interface Guidelines.
- [14] The J2EE 1.4 Tutorial; Eric Armstrong, Jennifer Ball, Stephanie Bodoff, Debbie Bode Carson, Ian Evans, Dale Green, Kim Haase, Eric Jendrock; March 31, 2004; Sun Microsystems, Inc.
- [15] Cascading Style Sheets, level 1; W3C Recommendation 17 Dec 1996, revised 11 Jan 1999; Hakon Wium Lie, Bert Boos; W3C.
- [16] HTML 4.01 Specification; W3C Recommendation 24 December 1999; Dave Raggett, Arnaud Le Hors, Lan Jacobs; W3C.
- [17] Web Server Administrators Guide to the Robots Exclusin Protocol.
<http://www.robotstxt.org>
- [18] Macintosh human Interface Guidelines; Apple Computer, Inc.; Addison-Wesley Publishing Company.