Maps of Information Spaces: Assessments from Astronomy

Philippe Poinçot (1)
Soizick Lesteven (1)
and Fionn Murtagh*(2)

(1) Observatoire Astronomique, 11 rue de l'Université, 67000 Strasbourg, France
(2) School of Computer Science, The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland

August 29, 1999

Abstract

We discuss the implementation of a cartographic user interface to bibliographic and other information subspaces in astronomy. This includes a front-end to two of the five premier scholarly journals in astronomy. We present a range of comparative assessments, in operational frameworks, of this approach to accessing and retrieving astronomical information. Finally we discuss the particular role that such cartographic user interfaces can play in Web-based information seeking, and contrast this with widely-used currently available search technologies.

Keywords: Concept spaces, Distributed information retrieval, Information

*Author for correspondence. Email f.murtagh@qub.ac.uk

seeking, Internet, Kohonen self-organizing feature maps, Maps, Neural networks, Resource discovery, User interfaces

1 Introduction

Information retrieval by means of "semantic road maps" was first detailed by Doyle (1961). The spatial metaphor is a very powerful one in human information processing. The spatial metaphor also lends itself well to modern distributed computing environments such as the World-Wide Web. Semantic road maps are not necessarily based on neural networks. In fact there are quite varied approaches to visual information retrieval interfaces (as can be verified by searching for "visual information retrieval interfaces" on the Web).

The Kohonen self-organizing feature map (SOM) method is an effective means towards this end of a visual user interface (Lin et al., 1991). As an alternative (or enhanced) approach, Zavrel (1996) favours an adaptive grid (growing cell structures network). Very close links between the SOM and a wide range of multivariate data display methods (Murtagh and Hernández-Pajares, 1995) similarly point in the direction of a large number of alternative design possibilities. As the last reference points out, however, the SOM method facilitates display on rectangular computer screen. The most extensive and influential use of the SOM user interface approach over the past few years has been around the work of Kohonen's team, where conference abstracts and newsgroup contents have been among the input datasets processed in this way (Kohonen, 1998; Honkela et al., 1998; WebSOM, 1998; Kaski et al., 1998). The application of SOM to Web-based text searching and retrieval is discussed in Chen et al. (1998), but as stated by those authors, the focus is on information of an "eclectic and non-cohesive nature" which differs from our objective in this paper. Lin (1996) has an interesting SOM representation of limited astronomical information from Yahoo and Excite.

The Kohonen SOM method is computationally tractable and produces stable results of good quality. Being a server-side system, it is compatible with widely-available software – for instance, no Web browser add-ons are required as is sometimes the case with other visual interfaces.

In this article, we are interested in the SOM method for operational use. Following a short description of various implementation details, we describe the setting in which it is used. A comparative assessment follows, based on an alternative well-established and very widely-used retrieval system. Results of retrieval experiments are then presented. In the Conclusion, we argue for the importance of information maps, and how they can help with the ever-growing and pressing need for tools to support information resource discovery and retrieval.

2 Background: Astronomy as an "All-Digital Science"

The quotation that "it is the destiny of astronomy to be the first all-digital science" (cited in Heck and Murtagh, 1993) is – with a pinch – an apt one. Much astronomical imaging begins with digital detectors. The data often end up in image archives, or as reduced (i.e. processed) data in catalogs. Data centers are needed, to support the organization of, and professional access to, data collections. Since the advent of the Web in early 1993 (we are dating this, of course, from the time of public release of the Mosaic browser), image and catalog data are ever increasingly available online for professional astronomers. Examples of data centers include the Space Telescope Science Institute (Baltimore, Maryland) which stores all Hubble Space Telescope data, with mirror sites in Europe and Canada; and the Strasbourg Data Center (CDS), Strasbourg Astronomical Observatory which evolved over recent decades to be a major center for catalogs. A wide range of data holdings are surveyed in Egret and Albrecht (1995).

Such data and associated information are characterized by growing size, growing complexity, but additionally by growing integration. Astronomy is fortunate in having its coordinate systems as an absolute information frame of reference. But there are often multiple nomenclature systems, and naming conventions come up against their limits when faced with problems of resolution and scale. While coordinate systems provide a very useful way to avoid problems with nomenclature, nonetheless the resolution dependence of measurement gives rise to other issues (see Read and Hapgood, 1992). In regard to such aspects, astronomy has unique information access and retrieval problems. However the underlying information infrastructure of astronomy is shared with many other domains and fields of study, and is due, ultimately, to the Web.

Alongside archives and databases, the astronomical literature is increasingly available online (Heck, 1997). The Astrophysics Data System (ADS), located at the Center for Astrophysics, Harvard, Massachussets, and with mirror sites in Europe, Japan and South America, is the best entry point to the astronomical literature (Murray et al., 1995; ADS, 1998). In astronomy and astrophysics, it contains about half a million abstracts. About 40,000 full journal articles are also available. It is with the very extensively-used ADS service that we carry out our comparative experiments on document maps, to validate our approach. We used keyword (controlled term) and free text access mechanisms supported by ADS.

The document and catalog maps used in this work can be accessed at the addresses listed under Poinçot (1998). The one related to the journal *Astronomy and Astrophysics* was used for the experiments reported on below.

3 Self-Organizing Map: Implementation

An SOM can be considered as a display grid where objects are classified (Figure 1). In such a grid, similar objects are located in the same area. In the example (Figure 1, right), a global classification is shown: the three different shapes are located in three different clusters. Furthermore the largest objects are located towards the center of the grid, and each cluster is ordered: the largest objects are at one side of a cluster, smaller shapes are at the other side.

An SOM is constructed as follows.

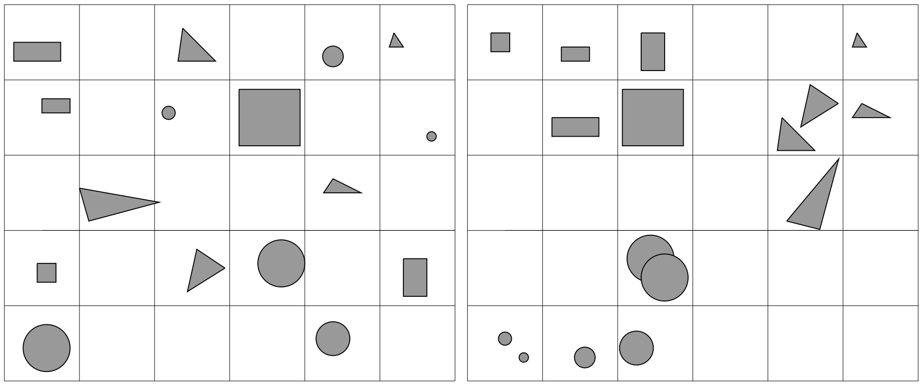

Figure 1: Object locations. Left: before learning. Right: after learning.

- Each object is described by a vector. In the example, the vector has two components: the first one corresponds to the number of angles, and the other one to the width of the area.
- Initially, a vector is randomly associated with each box (or node) of the grid.
- Each document is located in a box whose descriptive vector is the most similar to the object's vector.
- During an iterative learning process, the components of the nodes describing vectors are modified. The learning process produces the classification.

Typical use of the SOM in order to classify bibliographical data was as follows. Our set of documents relates to the journal *Astronomy and Astrophysics* in the time-interval 1994 to 1998. The descriptive vector is based on the journal keywords associated with each document. We eliminated rare keywords (found in less than 5 documents). Finally, we used about 6500 documents described by 269 keywords.

Adaptation of the SOM for our purposes included the following steps.

- Documents located at a map edge have neighbors at the other side of the map. It is then possible to reconfigure the map without losing the similarity of closely clustered documents (Figure 2, left).

- When there are many documents (more than 30) associated with a node of the map, we create a new map with the documents attached to this node and the 8 nodes around it. Such a map is called a *local* map or *detailed* map (Figure 2, right). The first map is called the *principal* map. We used a 15×15 grid for the principal map, and a 5×5 grid for the detailed maps.

We do not wish to describe our implementation further here, and instead refer the reader to Poinçot et al. (1998). Among the issues dealt with in detail in that paper are the following.

- Influence of the number of map nodes on the result, leading to implementation of primary and secondary maps.
- Estimates of time required to train the map.
- Weighting of index terms.
- General considerations on the construction of the density maps.
- Annotation of the maps.
- Creating the user interface to the maps.
- Finally, an in-depth example of map interpretation is provided.

Next we come to the graphical user interface. We display the trained SOM map as a density map, which represents graphically the areas containing papers of similar content and the number of documents in these areas. The map is labeled to locate on it the themes dealt with (Figure 3, top).

The user can select one node of the map (by clicking on the image) to obtain information about the articles located at this node (the number of documents and the keywords describing them appear on the right side of the interface) (Figure 3, bottom). The user can also access the detailed map, and/or the article content (title, authors, abstract) and all the facilities provided by the

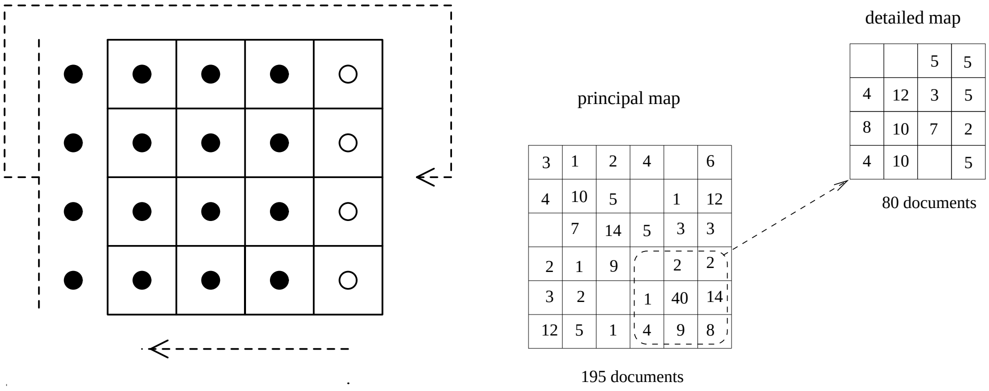

Figure 2: Our use of SOM. Left: shifting the rows or the columns. Right: two classification levels.

CDS bibliographical service (including a link to ADS and to the online full paper in many cases).

The user interface allows one to select and display on the map only a part of the database. This is used with keyword queries (only the documents containing selected keywords are shown), or with an external list of documents (bibcode queries: a bibcode is a 19-character standard for a document reference, an example of which will be seen at the beginning of section 4.1). We will refer to this latter form of querying in section 4.3.2 below.

Figure 4 gives an overall view of the system. It is based on imagemap and CGI scripts. Hence our implementation is server-side and compatible with any browser supporting imagemaps.

4 Comparative Assessment of the SOM Map

In order to validate our document retrieval tool, we initiated a study based on the comparison of our system with the ADS system, taken as a reference. The ADS provides wide query possibilities, so we were able to use exactly the same set of documents with the two systems.

Two different types of comparison were carried out. We first used the ADS

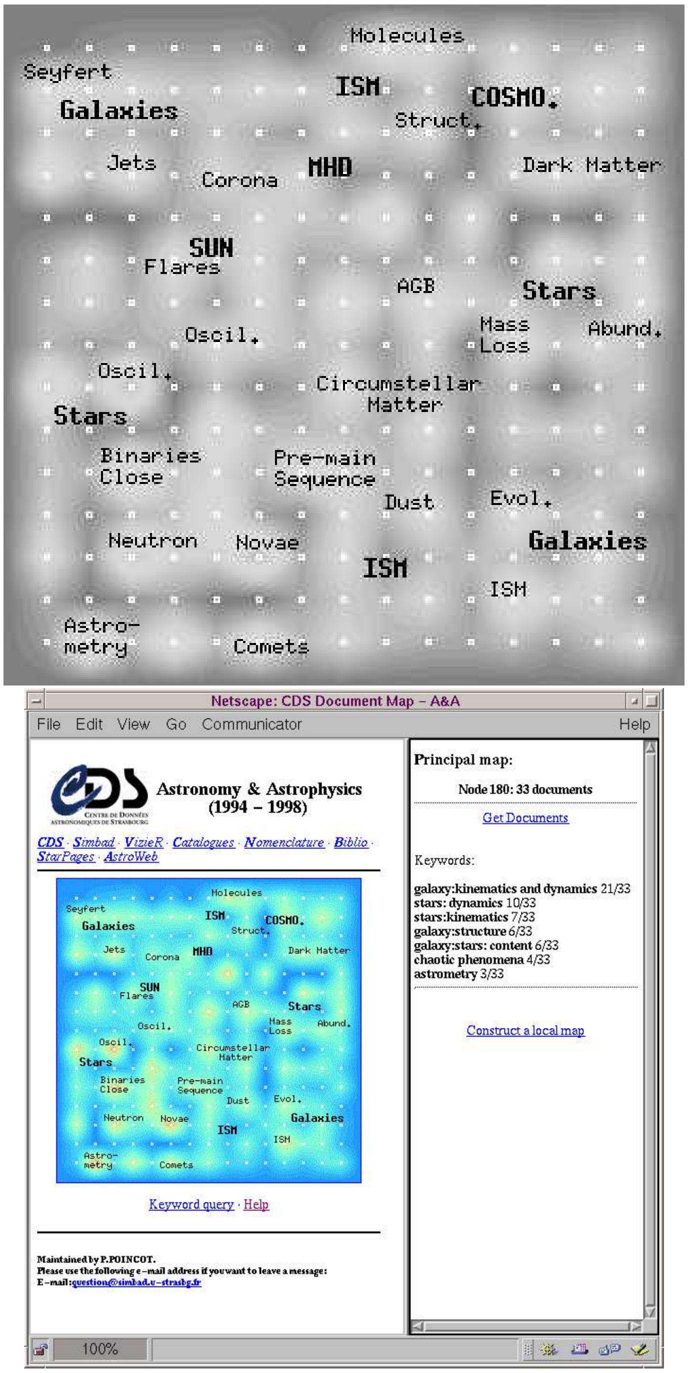

Figure 3: Top: the principal map. Bottom: the user interface. URL http://simbad.u-strasbg.fr/A+A/map.pl

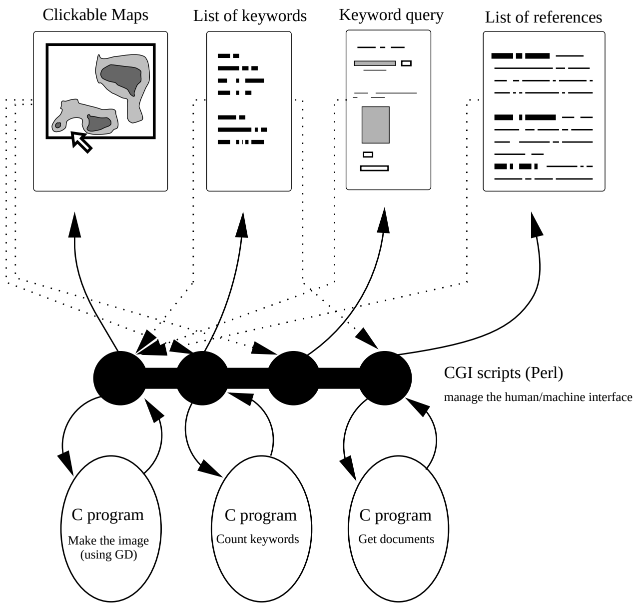

Figure 4: SOM-based system.

with a keyword query (i.e. using a controlled list) to validate more precisely our system. For the second type of comparison we used the ADS with a free text query to compare the two systems with similar criteria for document retrieval.

Differences in the two systems include:

- The bibliographic map is based on keywords only; an older version of ADS can handle keyword search or a free text search.
- The bibliographic map provides a graphical view of classified documents; ADS returns a sorted list of documents, corresponding to the user's query.

The comparative assessment methodology is as follows. We select one document, calling it the "starting document", and retrieve similar documents (from the same bibliographic database) independently using the two systems. Then, the different sets of documents are compared by an expert in the field.

More precisely:

- With the bibliographic map, we first locate the node containing the "starting document" (the winner node), and we get a first set (list 1) of documents with the contents of this node and the 8 surrounding nodes. A detailed map is built with these data, and then we get a second set of documents (list 2) containing the documents located in the detailed winner node and its surrounding nodes. The first list is much longer and maybe noisier than the second, but it allows more similar documents to be found.

- With the ADS, we used the "similar documents search" by keywords facility, where keywords describing the starting document are searched for:
  - in the document keyword field, giving us a first ADS list (list A)
  - in the document full text, giving us a second ADS list (list B)

  Then, we give the different sets of documents to a domain expert who works on the subject of the starting document. He/she gives a score for each document.

Eight domain experts contributed to these appraisals. They are staff members of Strasbourg Astronomical Observatory, interested in the map user interface, and prepared to assess it in their particular domain of interest. The "starting document" used in each case reflected this domain of interest. The domain experts were familiar with the very widely used ADS service, which would tend to more than balance any favoritism towards a locally produced system (which, moreover, they had no role nor material interest in developing).

4.1 Analysis using precision and recall

We first discuss a set of overall results and then look at one such case in greater detail.

Figure 6 shows a range of recall-precision plots using only one node of the principal map and the full text analysis of ADS. For each plot, we selected a starting document, we retrieved similar articles with ADS, and we retrieved articles located in the same node as that in which the starting document is located.

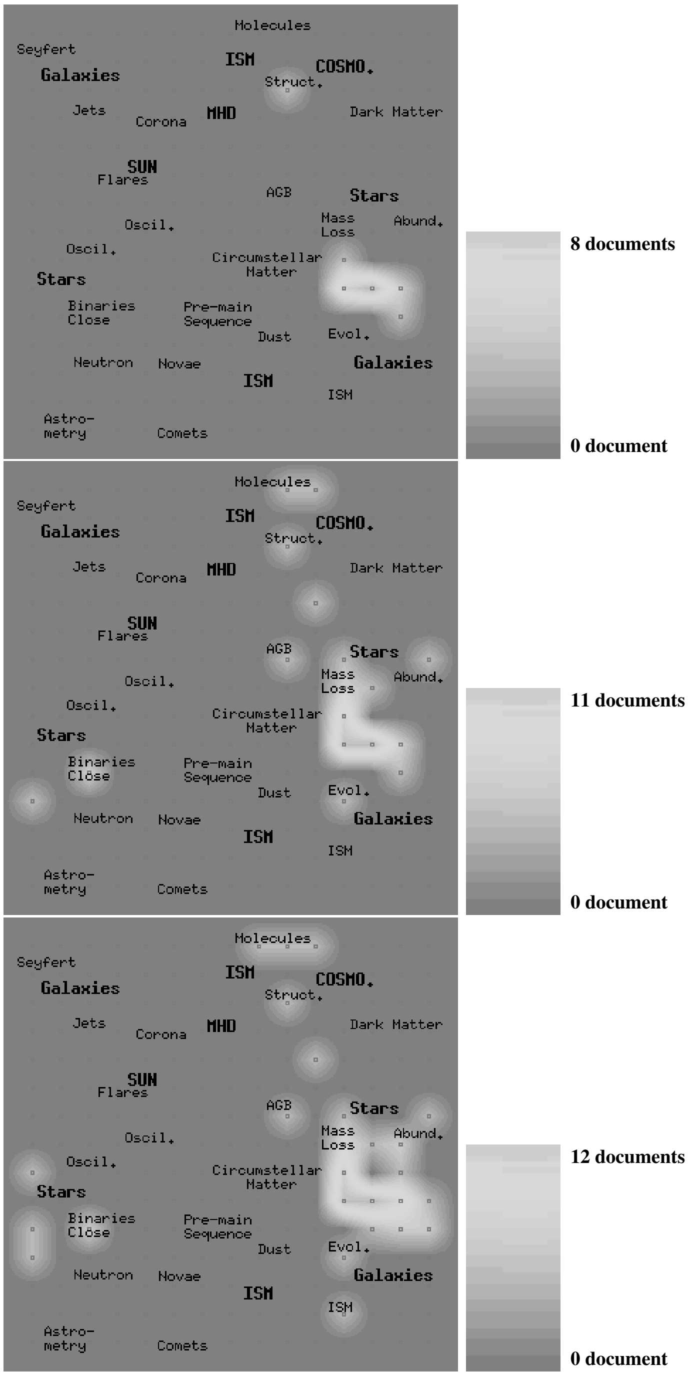

Figure 5: Spreading of documents retrieved by the ADS. Left to right: 20, 40 and 60 documents.

Each point on a curve corresponds to the addition of a document to the list of documents retrieved. When a curve descends, a document was obtained which was not relevant. An increase in the graph corresponds to the case of relevant documents. Some small aspects of this study are unfavorable to our system: in Figure 6, the documents were ordered by similarity with the node centroid, which can differ to a greater or lesser extent from the starting document. This especially happens with an uncommon starting document, or a document with a relatively uncommon theme. In such cases, the 20-odd documents retrieved are not explicitly ordered by relevance, which would have improved the appearance of the graphs in our favor in some cases. Therefore the order of introduction of documents retrieved by the map, whether relevant or non-relevant, is arbitrary in this Figure.

The nomenclature ("Exp1", "Exp2", etc.) represent the different experiments. The top left, bottom left, and bottom right, results are roughly similar. The top right and middle right results favor our map approach. The middle left plot favors ADS.

Figure 7 presents averaged curves. Our "Map" method is disadvantaged by the bad result in the third (counting from upper left) result in Figure 6. This bad result, in turn, is explained by the theme (brown dwarf stars), which we suggest was not as cohesive as other areas and hence less appropriate for clusterwise retrieval.

Again with reference to Figure 6, in the case of the fifth (bottom left) of these curves, we show example documents obtained in Table 1.

4.2 Detailed experiment

Clearly the different retrieval methods, our SOM-based one which we have called "Map", and the Astrophysical Data System, ADS, offer various information navigation possibilities. Having looked at a summary of comparative results, we will now consider these varying navigation possibilities in more detail based on a single experiment. This experiment is the last one shown (bottom right)

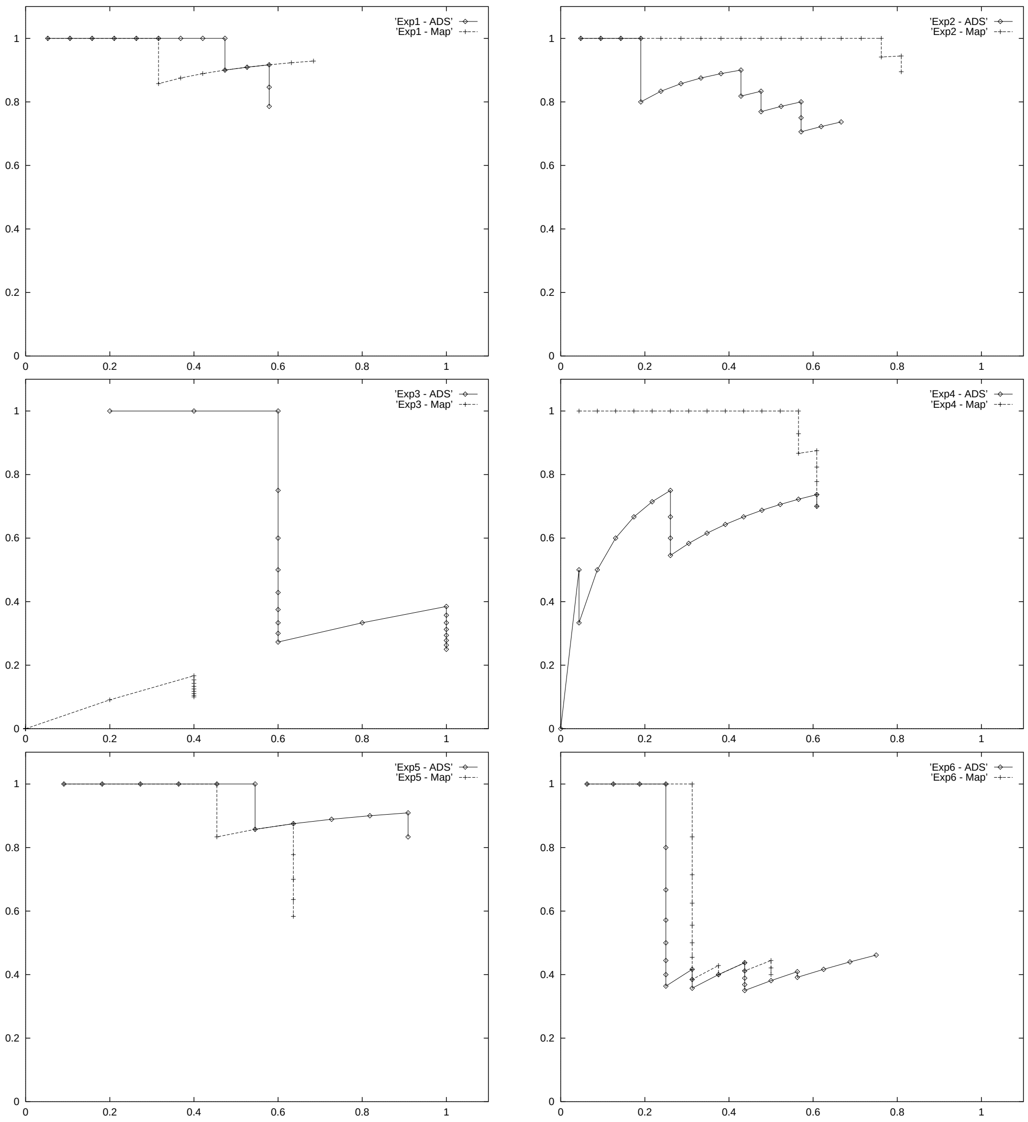

Figure 6: Six different experiments based on six different documents, showing recall (horizontal axis) and precision (vertical axis) for our method ("Map") and for ADS using keyword search. See text for further details.

Sarting document:

Carbon, nitrogen, oxygen and lithium abundances of six cool supergiants in the SMC. HILL, V.; BARBUY, B.; SPITE, M. Astronomy and Astrophysics, v.323, p.461-468, 1997

Relevant:

Two nitrogen rich main sequence B-stars in the Small Magellanic Cloud cluster, NGC 330. LENNON, D.J.; DUFTON, P.L.; MAZZALI, P.A.; PASIAN, F.; MARCONI, G. Astronomy and Astrophysics, v.314, p.243-250, 1996

Chemical evolution of the Magellanic Clouds. VI. Chemical composition of nine F supergiants from different regions of the large Magellanic Cloud. HILL, V.; ANDRIEVSKY, S.; SPITE, M. Astron. Astrophys. 293, 347-359 (1995), 1995

Non-relevant:

Fundamental parameters of Wolf-Rayet stars. VI. Large Magellanic Cloud WNL stars. CROWTHER, P.A.; SMITH, L.J. Astronomy and Astrophysics, v.320, p.500-524, 1997

Red supergiant variables in the Large Magellanic Cloud: Their evolution and pulsations. LI, Y.; GONG, Z. G. Astronomy and Astrophysics (ISSN 0004-6361), vol. 289, no. 2, p. 449-457, 1994

Table 1: Examples of relevant/non-relevant documents obtained.

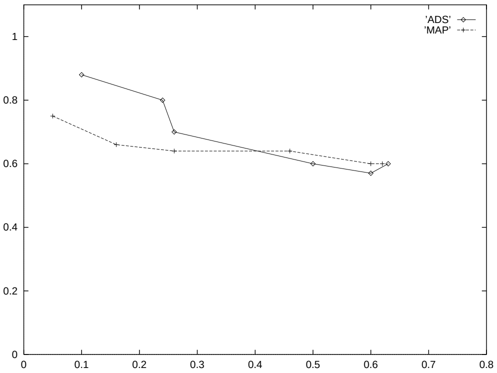

Figure 7: Averaged "Map" and ADS recall (horizontal) and precision (vertical) measures.

in Figure 6.

As a starting document, we used the following: "Metallicities and carbon abundances of 5 red supergiants of the SMC globular cluster NGC 330", Meliani, M.T. et al. with bibcode `1995A&A...300..349M` (i.e. published in *Astronomy and Astrophysics*, 1995, volume 300, start page 349, and first letter of first author name M). This article is described by the following keywords: GALAXIES: ABUNDANCES, MAGELLANIC CLOUDS, STAR CLUSTERS, GLOBULAR CLUSTERS: NGC 330 (SMC), STARS: ABUNDANCES.

4.3 Validation with the "keywords only" ADS version

4.3.1 Results from the CDS map

The starting document is located in node 6 of the detailed map, and only 2 of the 8 surrounding nodes contain documents (node numbers 5 and 11). See Figure 8. CDS list 2 contains 19 documents with scores given as shown in Table 2.

4.3.2 Results from "keywords only" ADS version

For the keyword query, the ADS retrieved 323 documents. The 315 last ones have the same small score because they are described only by one keyword of the query. We eliminated these documents for this reason. There remain 17

0 0 doc.	1 0 doc.	2 0 doc.	3	4
5 4 docs.	6 9 docs. (Winner)	7 0 doc.	8	9
10 0 doc.	11 6 docs.	12 0 doc.	13	14
15	16	17	18	19
20	21	22	23	24

Figure 8: CDS map result: illustration.

documents. Only one is scored as less relevant by the specialist.

4.3.3 Comparison

- 11 relevant documents are simultaneously retrieved by both systems.
- The bibliographic map retrieves 3 documents scored as relevant which ADS does not retrieve, but it misses 5 others. Three of these documents are located in the Inter Stellar Medium (ISM) region, and can be retrieved by a keyword query with the bibliographic map, using the keywords describing the starting document. The two others are located in another node of the detailed map.
- ADS retrieves 5 relevant documents (scored as relevant by the expert) that the CDS does not retrieve, but misses 3 other relevant documents. One of them is only described by one keyword of the starting document, the two others are described by another set of keywords (from NASA-STI, NASA's Scientific and Technical Information system, another controlled system).

These initial results show that both systems retrieve almost the same documents. They provide preliminary validation of the SOM method when used for the purpose of document retrieval.

node	relevant documents	less relevant documents
6*	8	1
5	4	0
11	2	4

*node containing the starting document.

Table 2: Results for the documents coming from the detailed map (list 2).

4.4 Validation with the ADS full text query

4.4.1 Results from the CDS map

We take into consideration all the documents of the detailed map, which contains 80 documents (list 1). A check by an expert shows that 22 of them are relevant, 58 are less relevant.

4.4.2 Results for the ADS full text query

To compare the second ADS list (list B) with the map, we selected the 80 first documents (ADS sorts the documents by decreasing similarity). As we have seen at the end of section 3, we can visualize on the density map an external list of documents when they are already classified on the map. In Figure 5, we can see the location on the map of the documents retrieved by ADS, corresponding to the given starting document. There is a spreading of the documents as the list becomes longer. This means that the smaller the score of a document, the further it is located from the starting document.

A check by an expert gives 25 relevant, and 55 less relevant, documents.

4.4.3 Comparison

- 15 relevant documents are simultaneously retrieved by both systems (Table 3).
- The bibliographic map misses 10 other relevant documents. These doc-

	relevant	less relevant
ADS (80 first documents)	25	55
map	22	58
common	15	22

Table 3: Results for all the documents coming from the detailed map (list 1), compared to the ADS (list B).

uments may be lost because associated descriptive keywords are wrong (typographic errors), or not accurate enough: i.e., relevant and less relevant articles are found to be described by the same set of keywords. Other lost documents may be retrieved with a keyword query (using the keywords describing the starting document), which shows articles in the ISM region for example (cf. Figure 5).

- ADS retrieves 10 relevant documents that the map does not retrieve, but it misses 7 other relevant documents. There are 7 relevant documents in the first 20 retrieved documents, 14 relevant documents among the subsequent 20 retrieved documents, and 25 in the whole set (list B).

The missing relevant ADS documents are retrieved if we examine the ADS result list beyond the first 80 retrieved documents. These results show that the list of documents retrieved by ADS (the first documents of list B) is noisier. But if we examine the complete list, ADS retrieves more relevant documents than our system.

Through ADS full text analysis, we can retrieve more relevant documents, but these are drowned in a longer set of documents.

4.5 Conclusions on the experimentation

This study shows that the bibliographic map gives results comparable with an ADS request by keywords.

When we compare our results to the ADS full text retrieval system, we show that ADS retrieves more relevant documents, but the bibliographic map is more accurate for the initial retrieved documents. Is seems reasonable that the full text search is noisier, but is more complete if full results are taken into account.

The two systems appear to be complementary, because the documents retrieved are about 60% in common, and both systems pull in other similar documents. It could be recommended to use both systems to get as much relevant information as possible.

Furthermore the bibliographic map, with its graphical interface, is very appropriate for use on the Web. While not the only such graphical interface (see, e.g., Lin, 1995), it has been realized recently is a large commercial system (Cartia, 1999). The keyword and bibcode queries give complementary access to the map. The fact that links exist between two neighboring nodes provides a new mechanism for helping users retrieve similar documents.

5 Conclusion: Maps in Resource Discovery

In a wide-ranging survey, White and McCain (1997) compare and contrast models – bibliographic, editorial, user, etc., and also visualization – for domain analysis and information retrieval. We do not emphasize the domain analysis aspect of our work: the Kohonen SOM map is a relatively blunt tool for this compared to the range of multivariate data analysis methodologies which have been refined over the decades. Regarding the information retrieval aspect, we found that it was useful to provide support also for keyword input by the user. Therefore the cartographic representation is not a well-rounded tool, either, for information retrieval. Where visualization work does score however is in providing a "responsive interface" (White and McCain, 1997). The major benefit of this work lies in the human-machine interface, – how people interact with machines for the quite specific objective of bibliographic resource discovery and information retrieval.

Retrieval of Web-based data and information is witnessing very great activity, with very wide use of a range of more or less successful search engines. Such search capability is based on indexing HTML and other documents.

The issues that we have addressed in this article are quite distinct from the issues and problems addressed by such search engine technologies. Increasingly the Web not only allows presentation of information, but also provides a common front-end to data and information which is not directly accessible in its entirety by a user. One can consider, in this context, database management systems which are frontended by a Web interface; or bibliographic data collections; or information served up by Active Server Page technology; or any number of other examples which come rapidly to mind. Traditional Web indexing and retrieval technologies crudely focus on the tip of the iceberg, and very limited amounts of the information iceberg adjoining the tip.

The problem of accessing the full information iceberg is not one that traditional technologies can handle. The disparate nature of what an information iceberg contains implies that indexing cannot work. Instead, the problem is firstly and foremostly a problem of describing the information iceberg, or information space. Giving the user an overview of the information space's contents is the first step. The user may then click further to enter into such an information space.

We have described a validated and operational solution to this problem. We see two new problems now arising. Firstly we should exploit the full generality of this work to cater for heterogeneous multimedia data. Secondly, we see the need, over the coming years, to define common standards for map representations of information subspaces. Our work will help towards these two goals.

References

[1] Astrophysics Data System, ADS. (1998). http://adsabs.harvard.edu

[2] Cartia, Inc. – Mapping the Information Landscape, software system, http://www.cartia.com

[3] Chen, H., Houston, A.L., Sewell, R.R. and Schatz, B.R. (1998). Internet browsing and searching: user evaluations of category map and concept space techniques, *Journal of the American Society for Information Science*, 49: 582–603.

[4] Doyle, L.B. (1961). Semantic road maps for literature searchers. *Journal of the ACM*, 8, 553–578.

[5] Egret, D. and Albrecht, M., Eds. (1995). *Information and On-Line Data in Astronomy.* Kluwer, Dordrecht.

[6] Heck, A. and Murtagh, F., Eds. (1993). *Intelligent Information Retrieval: The Case of Astronomy and Related Space Sciences.* Kluwer, Dordrecht.

[7] Heck, A., Ed. (1997). *Electronic Publishing for Physics and Astronomy.* Kluwer, Dordrecht.

[8] Honkela, T., Kaski, S., Kohonen, T. and Lagus, K. (1998). Self-organizing maps of very large document collections: justification for the WebSOM method, in I. Balderjahn, R. Mathar and M. Schader, Eds., *Classification, Data Analysis, and Data Highways*, Springer-Verlag, 245–252.

[9] Kaski, S., Honkela, T., Lagus, K. and Kohonen, T. (1998). WEBSOM – self-organizing maps of document collections. *Neurocomputing*, 21, 101–117.

[10] Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, 21, 1–6.

[11] Lin, X., Soergel, D. and Marchionini G. (1991). A self-organizing semantic map for information retrieval. In A. Bookstein, Y. Chiaramella, G. Salton and V.V. Raghavan, Eds., *SIGIR'91, Proceedings of the 14th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, special issue of *SIGIR Forum*, 262–269.

[12] Lin, X. (1995). Searching and browsing on map displays. Presentation, *ASIS'95 – American Society for Information Science*, http://www.uky.edu/~xlin/asis95.htm

[13] Lin, X. (1996). Space Science and Astronomy and Space maps, accessible from http://lislin.gws.uky.edu/Sitemap

[14] Murray, S.S., Eichhorn, G., Accomazzi, A., Stern Grant, C. and Kurtz, M.J. (1995). Data holdings in the ADS. In D. Egret and M. Albrecht, Eds., *Information and On-Line Data in Astronomy*, Kluwer, Dordrecht, 175–184.

[15] Murtagh, F. and Hernández-Pajares, M. (1995). The Kohonen self-organizing map method: an assessment, *Journal of Classification*, 12: 165–190.

[16] Poinçot, P. (1998). SOM maps. *Astronomy and Astrophysics* document map, http://simbad.u-strasbg.fr/A+A/map.pl. *Astrophysical Journal* document map, http://simbad.u-strasbg.fr/ApJ/map.pl. Catalog document map, accessible on page http://vizir.u-strasbg.fr/viz-bin/VizieR.

[17] Poinçot, P., Lesteven S. and Murtagh, F. (1998). A spatial user interface to the astronomical literature. *Astronomy and Astrophysics Supplement Series*, 130, 183–191.

[18] Read, B.J. and Hapgood, M.A. (1992). Approximate joins in scientific databases in practice. In H. Hinterberger and J.C. French, Eds., *Proc. Sixth International Working Conference on Scientific and Statistical Database Management*, ETH Zurich, pp. 123–131.

[19] WebSOM (1998). A novel SOM-based approach to free-text mining, Web site, http://websom.hut.fi, latest update 1998.

[20] White, H.D. and McCain, K.W. (1997). Visualization of literatures. In M.E. Williams, Ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. 32, 99–168.

[21] Zavrel, J. (1996). Neural network interfaces for information retrieval: are they more than an appealing idea? *Artificial Intelligence Review*, 10, 477–504.