

Methods for defining user groups and user-adjusted information
structures

Pedro Valero and Jaime Sanmartín

* University of Valencia, Spain.

Contact address:

Pedro Valero

Departamento de Metodología, Psicobiología y Psicología Social

Avda. Blasco Ibañez 21, CP: 46010

Valencia. España (Spain)

Telephon: +34(96) 3864420. Ext. 6290

Fax:+34(96) 3375971

Email: valerop@uv.es

Methods for defining user groups and user-adjusted information structures

PEDRO VALERO

JAIME SANMARTÍN

UNIVERSITAT DE VALÈNCIA

ABSTRACT

A common problem in the design of information systems is how to structure the information in a way that is most useful to different groups of users. This paper describes some statistical methods for revealing the structure inherent in empirical data elicited from users. It is illustrated by the application of these methods to the design of some web pages giving information about the Universitat de Valencia. Three potential user groups were identified, administrative staff, teaching staff and students. The first analysis demonstrated that users within these three groups assign relatively homogeneous structures, but that the structures assigned by the three groups are not the same, and also, teaching and administrative staff were shown to be relatively similar and different from students. Second, the ideal information structures for each group were identified and validated against the original data. The methods described can be applied to any design situation where there is an existing user population that can be called on to provide data.

1. Introduction

This work deals with methods for the organisation of the information normally found in computerised public access systems, including those that can be used through Internet. However, it is not restricted to just this type of system. These servers are normally built by or supervised by just one person or organisation and are not therefore the end result of an unplanned and/or undistributed development. They (the servers) are normally used sporadically by individuals who are not specially trained in their use. Servers can be strategically important for their creators as they are a means of transmitting services or an image and their design could affect the users negatively or positively. Taking the example of a University, which is the focus point of the empirical part of this work, this server could offer information on activities, studies, grants etc., which could be of interest to the student. Depending on the quality of the design, the information could either be obtained quickly and accurately by the student or, as the case may be, the student could simply waste his or her time battling with the system.

Although there are many potential problems associated with the user friendliness of this type of system, this work will concentrate exclusively on the structure of the information. This problem bears resemblance to the classification of objects according to their similarities when the dimensions that must be attended to have not been previously specified. Baylay (1994) gives the following as an example: three people are requested to classify a group of kitchen utensils. One of them could tell the difference between forks, knives and spoons. The second could differentiate between plastic, wood or metal cutlery. The third and last person could classify the utensils according to size (large, medium and small). In this example, each subject had produced a correct classification but had based it on completely different dimensions, giving us a

radically different outcome. In real-life situations, the problem would be probably even more complex because it would not be so easy to guess the spectrum of possible solutions that the users could provide. Fortunately, there are statistical analysis methods that are appropriate for these situations and they allow us to determine the categories of information and the dimensions used to obtain them. Benysh et al. (1993) exposes information-collecting methods apt to determine the structure of knowledge of the users. These methods (McDonald et al. 1988, Vora et al. 1994) have proven to be capable of revealing the information structure from the point of view of the user when applied to problems of Human-Computer Interaction (HCI) and appear to be well enough established to require further testing of their efficacy in this context.

How to apply these methods is not all clear particularly for a designer who is not familiar with these techniques. This paper indicates the right options for the problem in hand, providing a set of tools for the designer building information structures adjusted to the user.

The following work centres on complementary aspects of the use of these methods from the point of view of a designer interested in using empirical data from the users as an aid to building an information server. We believe that there are three main questions that any designer would wish to answer.

a) Are there groups of users in the population that maintain their own or identifiable information structures? Can we detect potential user groups in our population of users? What is the right number of user groups we should pay attention to?

Determining if different groups of users maintain different information structures is important because if this is the case then the designer should build a range of views onto the database attending to these differences. This is

supported by the relevance of knowledge structures on the execution of various tasks that have been proven by, for example, Koubek and Moontjoy (1991). They show that subjects having a more abstract structure of a computer system are capable of finishing a task with a word processor before those subjects that possess a more specific one.

The differences between novice and expert subjects are one of the most habitual in literature. Experts appear to distinguish themselves from novices not only in terms of the number of relationships between concepts but the former also use dimensions that are not accessible to those (Chi et al 1981). Other individual differences that appear to be important in HCI are (Egan 1988): a) experience with the system, something which can be confused with the distinction between experts and novices; b) age; c) specific knowledge of the area.

These wide user categories would probably have to be given more detail in each specific case. In the example used here there are three apparent user groups: Students (ST), Teaching Staff (TS) and personnel working in the information services (IS). The first two are receivers and should be capable of managing the system in such a way as to be able to retrieve the information they need. The third group are not only users but also in charge of building and maintaining the information server. Hence, the hypothesis to be tested is that, given that these three groups of users have a different knowledge of the working of the university, they will have different views about the structure of the information. Methods to test this hypothesis are available in the literature and will be described and applied to the case study along this paper.

An alternative problem is to examine the data to find out whether different organisations of groups of users not previously anticipated can be identified.

This is an exploratory analysis v. the confirmatory analysis previously described and can be useful when relationships not foreseen by the analyst are present in the data.

b) What information structures do the users have? If there are differences between groups, what are these differences? What is the ideal structure of information for each group of users?

If we can determine the information structure by using the users themselves, they might find it much easier to use the systems. However, as we mentioned before, it is very possible that the users are themselves individually different and may even disagree considerably when it comes to their conceptions of information. Therefore, it is equally important to find regularities in the answers that satisfy a large number of users. Sometimes however, we might find ourselves with a group of users that share the principal idea but have certain differences in other aspects. Knowing these differences can be very useful to the designer because it allows for the building of servers that adapt to the different types of users.

c) Is the structure of a server proposed by the designer different to that of the users? Does the designer have a structure of the information compatible with the different groups of users?

Be it because the information server has been built without the user participation or because technical or organisational pressure has led to erroneous solutions, the designer needs to know up to what point the proposed server differs from that held by the users so that if there appears to be too much incompatibility, some modifications could be made.

The statistical techniques available to answer a), b) and c) are relatively complex with many alternatives available. The person using them runs the risk of using an incorrect alternative and of thereby obtaining incorrect conclusions (Arabie 1993). The appropriate solutions are described in Section 3. Next, Section 2 will describe the design setting that these alternatives were applied to before starting with the main body of this work.

2) Step 1. Eliciting data structures from users.

La Universitat de València (Spain) recently started a process of updating and improving the information and services offered on the Internet. One of the problems detected was the lack of structure of the information server, which led to difficulties finding one's way around the site. This was fundamentally due to the absence of a global standpoint in the system. A multidisciplinary group formed by members of the university, including the authors of this work, began the task of correcting this situation and redefining the role of electronic systems in the distribution of information. In relation to this, our work was centred on the study of the needs of the users.

Our first efforts to try to generate a more coherent structure were done by studying those employees who were in charge of distributing the information. For this we used a series of methods which will be described in the next section. We obtained an information structure for each of these workers (Valero et al. 1996). These structures were combined to obtain a general alternative that would become one possible proposal for our information server. However, we did not feel that the results obtained were satisfactory. The developed structure gave us the impression of being determined by administrative questions that other members of the university were not aware of. For example, some of the

classifications produced were justified by “ this goes in the same place because X department is in charge of it...” and similar types of argument. When we asked the participant subjects to define the tasks undertaken by X department, they could do so in a circular way referring to the items that they had previously marked as being part of it, or more honestly, confessing that it was very difficult to explain the logic behind all the tasks that a particular service was in charge of but that it was habit or tradition that had established things in this way. This was not a problem for the users that had enough time to live with this system of working. However, there are many other users, above all students, who do not know the details of the bureaucratic layout and may therefore find it rather problematic to navigate through it. Something similar could occur with the teachers because, even though they know the university organisation better, their knowledge is reduced in most cases to specific aspects with which they have a more habitual contact.

2.1. Goals.

The goals of the following study were to obtain a more profound knowledge and understanding of the way in which the future users of the computerised service would organise their information, as well as obtaining a degree of sufficient evidence to be able to influence the other members of the team involved in its development. In particular, we were interested in answering the three questions mentioned previously about whether or not different users maintain different information structures, as in this structure, and if the shape of the design would be compatible to all the users.

2.2. Subjects.

Twenty subjects participated in the study. Of these, six were all the members of the Student Information Service (IS). The size of the other groups was chosen in such a way that they did not differ too much from this figure. In this way answers were obtained from six members of the teaching staff (TS) and from eight students (ST).

2.3. Material.

For the study we selected information items that we considered typical in relation to the university community in general. At the same time we tried not to give too much representation to specific themes or contents etc. of any group. These were extracted from diverse sources such as the computerised information server that existed before beginning this project, posters, notice boards, internal memos, written material and the like. We selected a total of 82 stimulus. Although this is a relatively high number for a study of this type, it is still quite small compared to the number of stimulus normally available to the university community. However, this sample was considered to be sufficient to establish the widest categories of information at our university.

2.4. Procedure.

There are various information-collecting methods that would be apt for the elaboration of knowledge structures as Benysh et al. (1993) mentions. Among them, and although in our research we allowed for other antecedents, we limited ourselves to those producing quantitative information. With this aim there are a variety of procedures (Davison, 1983): direct similarity judgement, graphical methods, sorting by category or magnitude estimation. All these methods have the common denominator of obtaining distance matrices or non-

similarities between stimulus, so those with high values imply objects with fewer common attributes.

The subjects were asked to categorise the information items in such a way that they could be related hierarchically.

The subjects received cards with a one-line description of each item. Their task was to organise the items into categories or groups of items that, according to them, were related and therefore should be found in the same place in an information system. As we were interested in finding a hierarchical structure, we asked them to group the categories into higher order categories until they reached a unique superior level. The items to be classified could be situated in any of the levels. An item like 'telephone number of the department director', for example, could be in the same branch but on a lower level to that of the 'list of staff of the university'. Lastly, they were allowed to include the same item in two different categories which would show the existence of a direct connection between this item and another category which in turn symbolises a direct connection that breaks the hierarchical structure.

The aim of the study was explained to the subjects and they were allowed to examine all the items before beginning the classification. They were also allowed to change their minds as they worked and only finished when they were happy with the final product. The task took approximately one hour for all the subjects.

To determine the distances between each item the criteria established were the following:

(See figure 1.)

1) The distance between an item and itself was zero, i.e., the diagonal in the example matrix will always be 0.

2) The distance between one item and another situated in the same category was zero. For example, $d(a,b)=0$.

3) The distance between one item and another situated in the category connected to an inferior or superior level was one. For example, $d(a,d)=1$

4) The distance between one item and another situated in another category of the same level was two. This value corresponds to a displacement to a superior level and later to a lower level, in the way that a hierarchical system of information would oblige us to do so. For example, $d(b,c)=2$.

5) The distance between two items situated in different places of the hierarchy but still directly connected would be one. For example, $d(c,e)=1$. These links are not necessarily symmetric so you might be able to traverse easily from one point to the another but the return could be difficult. This is simbolized in figure 1 with a one-way arrow. An example is $d(e,c)=3$, so it easier going from c to e than from e to c. The matrix of distances so obtained is asymmetric.

****FIGURE 1 ABOUT HERE****

In general, distances are determined considering the number of steps necessary to reach one point from another using the hierarchy. The basic result can be thought as a ordinal scale where bigger numbers mean bigger distance between items.

A computer program was developed that took as input a file with the aforementioned information and produced as output the matrix of distances between items for each subject (available on request from the first author).

3) Step 2. Applying the statistical techniques.

This part provides an explanation of the methods to be applied and constitutes the main part of this paper. The university information server is both an example and a way of validating our approach to the questions raised in the introduction. Each question is stated first, then the methods for dealing with it will be described and finally the results from the example will be shown. A short conclusion ends the paper.

a) Are there groups of users in the population that maintain their own or identifiable information structures?

Determining if the structures generated by the different groups of subjects are similar is of great practical importance. If this were not the case, the designer would find him or herself before a series of arguments in favour of building different information servers for each one of them. The methods shown here allow us to approach two related but nevertheless different goals: 1) propose and respond to specific hypotheses about the possible groups of users and 2) explore the similarities between objects to determine if there are groups that were not anticipated.

The first step is to obtain an index of similarity between the subjects. To do this, the statistic r (see formula 2) can be applied to matrices for each pair of subjects (Hubert 1979). This produces the same result as computing a Pearson correlation coefficient between the values of two matrices after excluding the

diagonal (Legendre and Vaudor 1991). The results are organized in a new matrix like in figure 2 and are the first step to use procedures for analysis of variance on measures of proximity between subjects (Hubert et al 1982).

****FIGURE 2 ABOUT HERE****

Values along the diagonal in this matrix corresponds to correlations between members of the same group. For example, up to down the areas along the diagonal would indicate correlations respectively among Students (ST), Teaching Staff (TS) and personnel working in the information services (IS). The rest would correspond to the pairs coming from different groups. In our case we would generally expect that the members of a group would correlate more among themselves than among members of other groups (i.e. the correlation's en zones along the diagonal would be generally higher than the rest). This situation would indicate the existence of homogeneity intra groups (defined as higher correlations between members of a group than correlations between members of different groups).

If there were homogeneity, we should be careful mixing the proposals from those subjects belonging to different groups. If, however, there were very little homogeneity between groups (defined as correlations between members of different groups similar or lower to the intra group correlation) we can on the other hand, ignore the division between groups and act as if they didn't exist. The best strategies to make confirmatory analysis in this situation appear to be based on the test of Mantel (Arabie and Hubert 1992, Hubert et al 1982, Legendre and Fortin 1989, Legendre and Vaudor 1991). This test has the advantage of avoiding the problem of the non-independence between correlations (Rounds et al 1992) that other methods that have been used in

published literature do not really control (Cooke and Schvaneveldt 1988, Gillan et al. 1992). The object of this test is to compare two matrices, one of which, can be specified as a matrix of hypothesis formed by ones and zeros. In our case, the matrix of hypothesis would have the same form as that of the previous figure shown but with a value of 1 in the triangles along the diagonal (correlations between members of the same group), and 0 in the rest. The other matrix would have (in our case) the matrix of proximities between pairs of subjects as in figure 2. To obtain the test value, the products by pairs of both matrices are summed using the following formula:

$$(1) z = \sum_i \sum_j x_{ij} y_{ij}$$

Where x_{ij} is the correlation between a given pair of subjects and y_{ij} is one or zero depending on whether it is hypothesised they are members of the same group of users.

High values of z , positive or negative, will indicate a coincidence between the labelled values such as one and the high values of the same sign be it positive or negative in that matrix that is to be evaluated. In our case the matrix is formed by ones in the positions corresponding to the correlation's between pairs of the same groups and zeros of the pairs of different groups. This permits us to calculate the equivalent of an analysis of variance over the matrix of proximity (Hubert et al. 1982).

Given that the sample distribution of the indices is not known, the evaluation of z is obtained by means of a permutation strategy, shuffling the rows and columns of the matrices and recalculating z a number of times. In the case study, 999 permutations of the original matrix were used. The level of signification is simply the number of times that the result of the permutation is above or below the z value reached originally divided by 1000 (Legendre and

Vaudor 1991). It is possible to calculate a standardised z value that we will call r (2), which is adjusted to the limits of 1 and -1. This is given by:

$$(2) r = [1/(n-1)] \sum_i \sum_j [(x_{ij} - \bar{x}) / s_x] [(y_{ij} - \bar{y}) / s_y]$$

Where x_{ij} and y_{ij} is the same as above, \bar{x} and \bar{y} are the average respectively of the X and Y matrices. S_x and S_y are the standard deviations for those matrices and n is the number of entries in any of them.

In the case study the result obtained from the Mantel test was $r=0.25413$ ($p=0.001$). With this result we can state that the intragroup agreement is greater than that of the intergroup agreement and there is therefore homogeneity in the groups in question at certain extent. Subjects from the same group show greater correlation among themselves in general than those subjects belonging to other groups.

Once it has been determined that there is enough agreement between the groups, the next logical step would be to check which groups account for this homogeneity. This is important because if only one of the groups are consistent in this aspect, it would be enough to create different servers for them while the non-homogenous groups could be mixed.

These comparisons were done by applying the Mantel test to each pair of groups. In each case a comparison matrix was created following the structure of figure 2 but for each pair (figure 3), and another of correlation's that would include only the subjects belonging to the pair in question.

****FIGURE 3 ABOUT HERE****

The result of the comparisons between each group is shown in table 1. This table has the same shape as figure 2 to make easier its understanding. As can be seen, the comparisons between Students and the other groups produces significant results (Teaching Staff v. Students $r=0.31$, $p=0.005$; Information Service v. Students $r= 0.34$, $p=0.003$). The difference between the other two groups did not reach statistical significance ($r=0.06$, $p=0.20$).

****TABLE 1 ABOUT HERE****

In the diagonal of the table there are mean correlation for each one of the two groups (obtained by adding the individual correlation and dividing them by the number of corresponding pairs) labelled $\bar{r}_{st\ st}$, $\bar{r}_{ts\ ts}$ and $\bar{r}_{is\ is}$. In the low part of the areas where different groups cross you can find the mean correlation between members of different groups, labelled $\bar{r}_{st\ ts}$, $\bar{r}_{st\ is}$ and $\bar{r}_{is\ ts}$. These values can be used indicatively in order to appreciate the meaning of the previous comparisons.

Taken together, the results indicate that the students have a clearer intragroup homogeneity with respect to the solution presented than the rest of groups and that this does not appear to be compatible with the members of other groups. On the other hand the members of the teaching staff and the student information services did not show any significant intragroup homogeneity with respect to each other which leads us to believe that it is reasonable to combine their solutions and make up one single group. This group will be defined as Teaching Staff-Information Services from this point onward.

A comparison between members of the Student group and the other two were also be made. The result of this test is $r=0.12$ ($p=0.04$), confirming there are indeed differences between the students and the other groups.

**** FIGURE 4 ABOUT HERE****

The previous test does not inform about individuals. For example, one student can have an structure of the information very similar to the structure of the Teaching Staff, or just the opposite, a teacher can have a view of the information very similar to that of the Students. This can also happen with sub-groups of subjects. For example, half of the teachers can be similar to Students and the other half to the members of the Service of Information.

In figure 4, we show the results of the hierarchical cluster analyses computed over the matrix of distances between subjects with the option of using average distance between clusters as unifying criteria. This analysis will allow us to identify groups of subjects according to patterns not determined previously in order to evaluate the results obtained in a more exploratory way.

The dendogram shows a first group that corresponds principally to Students (symbol •) and only two subjects belong to the other two groups (* y – indicate Teaching Staff and Information Service respectively). It is possible to determine a second cluster composed of members of the other two groups, and, finally three individuals that appear to show very different solutions. This result confirms that the fundamental distinction can be found between students against members of the other two groups and no further subdivision is needed. We will continue with this distinction from here on.

b) What information structures do the users have? If there are differences between groups, what are these differences? What is the ideal structure of information for each group of users?

Once we have determined the interest to study the representations of at least two groups of users, the following step is to discover exactly what this representation consists in. This would allow us to determine the structure of the information that a computerised system would reproduce. The different proximity matrices between objects produced by the subjects involved in our study are the raw materials for this analysis.

There are two dominant approaches when it comes to the representation of proximity data (Klauer 1989): the spatial approximation (Young 1987) and the theoretic-graphs approach (Carroll 1995, Schvaneveldt 1988, Hutchinson 1983, Klauer 1989)

The methods that produce a spatial representation using distance data are known as Multidimensional Scaling techniques (MDS). Their purpose is to get a representation in a reduced dimensionality space of the objects that have information on the degree in which they are related. The distances between the objects in the space defined by the new dimensions try to reproduce the original information about the relationship between these objects. This result is interesting from a theoretical point of view when the dimensions reproduced are very similar to that hypothesised by the researcher. It is just as interesting from a practical viewpoint. It allows one to observe the objects by means of a more comprehensible representation. One family of MDS models is especially interesting for the case study. These are known as Multidimensional Scaling of Individualised Differences (INDSCAL) and are the answer to the problem of obtaining a spatial representation using various matrices that correspond to

different subjects or experimental conditions. Young (1987) comments that without INDSCAL, the researcher has two options: "...either average overall the matrices to obtain a single matrix (thus getting a headache from those who tell him he should not average) or must analyse each matrix separately (thus getting a headache from trying to interpret all the separate results)" (p27).

INDSCAL can be described as a weighted MDS model. In it we can obtain: (i) a stimulus space similar to that obtained in the non-weighted models that represents the information that is shared by all the individuals about the structure of the stimulus, and (ii) a space of individual weightings for these dimensions that allows us to see how the different subjects use more or less a given dimension when it comes to establishing relationships between the stimuli. In this way, the dimensions determined in the analysis, combined with the weighting of each subject, allows us to recuperate the individualised representation for each subject (Young 1987). However, the spatial representation is not adequate for our aim of building an information system because it does not fit to the technical characteristics of the software used to implement in the way that a hierarchical representation or a quasi-hierarchical would.

Klauer (1989) distinguishes between two types of graph-theoretical representations. In the most traditional, the representations are limited to a special one in which cycles are not allowed. There are in fact trees that implicate nested or hierarchical structures. These representations have the advantage of being clearly interpretable because they offer non-overlapping divisions of the block of objects. However, there are data structures that can be better described by other more general graphs such as networks or cycles that are not detected if methods limited to trees are used. Both structures offer a certain interest in our case. Even though electronic systems that allow for the

implementation of this type of system are usually described as following a network structure or 'web', these webs are in fact produced by the mere intersection of diverse sections built separately. Taken individually, these sections usually have a quasi-hierarchical or hierarchical structure. This structure normally consists of an initial menu that allows for the distribution of different contents through which the user must 'ascend' and descend' to explore. It is sometimes possible to have a direct or transversal path but this is usually secondary, the hierarchical structure being the most prevalent.

A method for obtaining network structures which has had some prevalence in the HCI literature is Pathfinder analysis (Schvaneveldt 1990, Schvaneveldt Dearholt and Durso 1988), a brief description of which follows: the Pathfinder algorithm takes proximity matrices and produces a network in which concepts are represented as nodes and relations as links. Concepts in the network could be directly linked or indirectly (so the path between two nodes requires a third node). A link remains in the network only if the direct link is shorter than any indirect link between the two nodes considered. A complete explanation of this method can be found in Schvaneveldt (1990). See Arabie (1993) for a criticism of the method and the subsequent responses, as well as Carroll (1995) for a partial support of the technique.

Determining if the subjects maintain a hierarchical or non-hierarchical structure of the information can be of great importance. A technique that produces hierarchical results is that of hierarchical cluster analysis.

On the other hand, Pathfinder obtains network results and can be useful when diagnosing up to what point a non-hierarchical result would be necessary for the problem in question.

The method that we propose is as follows.

1. Apply an INDSCAL model for each of the groups determined in the previous section with the aim of determining a global solution for each one of them.

2. Calculate the Euclidean distances between the objects for this solution

3. Obtain a Pathfinder analysis and hierarchical clusters on the distances from step 2.

4. Diagnose up to what point it is possible to use a hierarchical solution and, if it this is possible, interpret the cluster analysis complementing it with the Pathfinder solution.

The results from the cluster analysis and from Pathfinder for the case study can be found in annex II. The list of items used can be found in annex I. In the following section we will comment on the results obtained for Students and Teaching Staff-Information Services.

a) Students

After studying the solutions with dimensions between 2 and 6, obtained using INDSCAL, we opted for the use of that with biggest dimensionality. This was based on the fact that although the interpretability shown was similar to both solutions, it was observed that there was a tendency to obtain finer and more precise categories as the dimensionality increased.

The dendrogram of the cluster analysis of the Euclidean distances between objects just as they were collocated using INDSCAL is shown in annex II. In it we can see the following nine clusters (the numbers in which the cluster can be found on the graph are indicated by brackets): 1) Telephones and addresses (1-39), 2) Student by-laws (9-74), 3) Centres listing, freshman day and student organisations (14-42), 4) School insurance (7-15), 5) Congresses and pedagogic adaptation course (48-62), 6) Awards and research related themes (57-44), 7) Grants and discounts (43-69), 8) European interchange programs and PhD

courses (31-20) and 9) Postgraduate Courses, masters, internships and job opportunities (60-67).

Careful observation of these clusters produces additional observations. The following is a list of what we consider to be the most interesting ones:

1) Cluster number 3 appears to correspond to initial information for a student who is beginning university life. In it we can find accommodation information (70), freshman's party (81) (one of the most important social events) and a list of studies and options that make up a type of initial orientation for the beginning of the academic year.

2) School insurance (items 7-15) appear as an individual cluster, away from other matters.

3) Everything to do with economic affairs appears in the same place (items 43-69). This includes information about subventions for national project (45) such as European grants (50) and luncheon vouchers (69). All possible sources of income are included in this area despite the differences in goals, amount of money and other considerations involved.

4) European programs (items 31-56) appear beside the doctorate courses (item 19 to 20).

5) Master courses, Postgraduate courses and training (computers, foreign languages etc.) appear alongside internships and job opportunities.

The pathfinder solution shows that a hierarchical structure appears to adapt itself correctly to the data obtained as there are no cycles or connections

between branches. The pathfinder solution also adds information about the items that are responsible for the union between clusters. This information can be especially interesting in a computerised information system in which certain items would be directly linked. Some of the most important links are the following:

1) Items 69 and 70 join the discounts to study English with job opportunities abroad. The connection is clear.

2) Items 37 and 73. Computer courses connected to internships and job opportunities. Maybe the students perceive the knowledge of computing to be a pre-requirement for getting a job.

3) The items 33, 31 and 32 are separated from item 73. Although all four of them are European student interchange programs, only the last one has the merit of internship and this separates it from the rest.

4) The students see school insurance (items 7, 8 and 15) directly connected to questions related to inscription to university (items 29 and 30) given that the payment of this insurance occurs when the student is enrolled.

b) Teaching Staff – Information Services.

Again, after studying the solutions with dimensions between 2 and 6, it was found that the solution with biggest dimensionality presented a better interpretability and for this reason it was selected.

The hierarchical cluster analysis of the interobject Euclidean distances determined by means of INDSCAL for the Teaching Staff-Information Service

group produced the following clusters: 1) Interchanges, grants and activities abroad (31-74); 2) Research, including research awards and project proposals (items 53 to 46); 3) PhD courses research grants and third cycle; 4) Student by-laws (items 26-10); 5) Listings, addresses and telephones (items 1 to 36); 6) Training courses, masters and post grad courses, public exams and congresses (items 37 to 67); and 7) Help/Aid information, which includes student activities, school insurance, travel schemes, etc. (79-64).

It is interesting to take note of the following:

a) The area covering European questions is a well-defined one with very precise sections and sub-sections.

b) There are rules and normatives that appear beside the subjects to which they refer. Item 17 (post grade normatives) and 27 (normatives and grants).

With respect to the results found by Pathfinder, is interesting to see the centrality of the areas related to grants that are later converted into union points among a great variety of other subjects. The corresponding items are 5, 27, 6, 50, 4 among others and they convert themselves in a node that unites questions referring to investigation (the branch starting in 58 and ending in 75), European programs (32, 32, 33), PhD courses (from 20 to 21), and even insurance information (15, 8, 7 and others). The consolidating role that the area of grants has does not appear in the hierarchical cluster.

c) Comparison between the groups.

One of the most interesting parts of this study is the comparison between the two classifications. We will comment some of the most outstanding differences.

The Student group unites all the economic questions, be they large quantities of money like grants or other types of economic benefit such as luncheon vouchers. The Teaching Staff-Information Service group however distinguishes between the different subjects to whom the money is directed and they therefore distribute it across the scheme.

The Student group creates a large information group referred to training and in it they include European interchange programs and PhD courses to which they later add all other courses in general as well as things related to internships and job possibilities. The Teaching Staff-Information Service group however clearly separate areas related to PhD courses and all these training possibilities.

The Student group sees a connection between school insurance and the enrolment processes while the Teaching Staff-Information Service group see school insurance related to the service of student information (the body traditionally involved in providing this information).

In some way, the student group appears to join everything related to the enrolment process. This process is infamous for its complexity and it therefore implies formal questions (forms, study lists etc.), as well as informal questions (extracurricular activities to be done throughout the year, freshman's party) and temporary situations (inscription periods). This notion appears above all in the Pathfinder solution for the students. This association of information does not appear in the case of the other group of users.

The Student group appears to be fairly interested in having a cluster of information on the job situation or professional internships. This grouping is not at all important for the Teaching Staff-Information Service group.

In general, we can qualify the Teaching Staff-Information Service solution as being formally the most correct and that of the Student group as being more opportunist and centred on its own interests. This contraposition is a challenge for the design process and it could be resolved by finding formulas that are compatible to both groups of users or indeed develop different servers that adjust to each structure.

c) Is the structure of a server proposed by the designer different to that of the users? Does the designer have a structure of the information compatible with the different groups of users?

Normally, and even in the case of not having undergone studies such as that being proposed here, the designer will have to make some compromises between considerations coming from other sources. Ideally this should not mean the loss of the compatibility with the structures maintained by the users. Assessing this compatibility is therefore quite interesting.

In our case and as part of a more extensive project into which this particular work was inserted, one of the members built an information server prototype. This came from a requirements analysis which was done by means of interviewing individuals that are key to distributing information in the university. As well as this, he had access to the partial results of this work (which may or may not have influenced him)

To undertake the comparison between the designed server by this member and the schemes put forward by the users, a matrix of distances was obtained between items, similar to that which we had built with the classifications made by the users. This was done by assigning the 82 items to their corresponding sections in the prototype. That was followed by the application of the described metric to the classifications of the users and exemplified in figure 1. A matrix with the distance between items was obtained. This matrix will be called Prototype Matrix from here onward in the text.

This matrix of distances was compared using the Mantel test with each of the matrix obtained through INDSCAL. In this case the test is different to that mentioned in the previous section given that none of the matrix are formed by ones and zeros. There is no hypothesis matrix and the comparison, when it is significant, implies that high values in one distance matrix have corresponding high values in the other distance matrix. In our case it was interesting to make the following comparisons: Prototype Matrix v. Students and Prototype Matrix v. Teaching Staff-Information Service.

The results had a p value higher than 0.01 significance in both cases. We have therefore Prototype Matrix v. Students r was equal to 0.27 ($p=0.00100$) and for Prototype Matrix v. Teaching Staff-Information Service r was equal to 0.40 ($p=0.001$). 999 permutations of the original matrix were used to obtain the probabilities.

The previous result indicates that Prototype Matrix is compatible with the solutions given by the different groups identified previously in this work. It is interesting to note that the correlation of Prototype Matrix with Teaching Staff-Information Service is greater than that with Students. This is easily explained

if we take into account that the former were consulted on the elaboration of the prototype while the latter were not.

We can conclude that the Prototype Matrix seem to be an adequate compromise between the models supported by the main groups identified previously. This test is consequently an indication of the possibility of avoiding the building of different servers of information for each group of users by the design of an "average" structure, able to satisfy different views of the database.

4) Conclusions.

The results of this study show the value of classification data obtained from users in the context of the development of the structure of information servers. This data, as has been seen, allows us to determine if there are indeed different user groups, what structure the information has in each case and, finally, to make comparisons with prototypes using a series of statistical and representational techniques. In fact, it is our opinion that the most interesting aspects of this work lies in the possibility of offering an analysis scheme that makes it easier for other researchers to approach this type of data in a similar situation. Obviously, our pretension is not to establish that only the techniques laid out here are useful in these situations. However, we consider that the applications of the Mantel test, along with the possibilities of the representation techniques of the proximity data, spatial and non-spatial, and, within these, hierarchical and non-hierarchical, suppose an analytical arsenal which should be strong enough to satisfy a great number of applied researchers interested in answering similar practical questions such as those set out in this work.

We would not like to end this work without making reference to some of the substantive results found in our example. When we began this study, our

original idea was that Students, Teaching-Staff and Information Services would maintain different information structures. Our results, however, lead us to believe that the truly important distinction is between Students and Teaching Staff-Information Services groups, which is similar to the novice-expert distinction. However, this work gives a slightly different hue to that distinction. Up to now, the expert's solutions are generally considered as being the best ones. In our case this would be just the opposite. The novices provided us with simpler and more homogenous classifications from the point of view of the applications that are considered here. The solutions produced by the experts are therefore much more heterogeneous. So, it seems that a deeper knowledge and understanding of the aspects of the working of the university leads to a lack of agreement between the different proposals of people about it.

This fact would not be so important if it were not for the tendency that we found throughout our study: i.e. to regard the 'expert's' information as the right one and consider the differences between them and that of the novices as a by-product of their 'ignorance'. Something which, although true, is converted into a difficult paradox to solve if we consider that in our case the novices were not only our 'clients', they also outnumbered the experts 40-1.

Acknowledgements

The authors would like to thank Dr. Forrest Young and Dr. C. M. Constanzo for help with questions about the statistical techniques used in this paper. We also would like to thank Dr. Andrew Monk for his many helpful comments and suggestions. Also, two masked reviewers contributed with several useful recommendations. This paper was supported by a grant from the

program for Scientific Research and Technological Development "Generalitat Valenciana".

Annex I

List of items

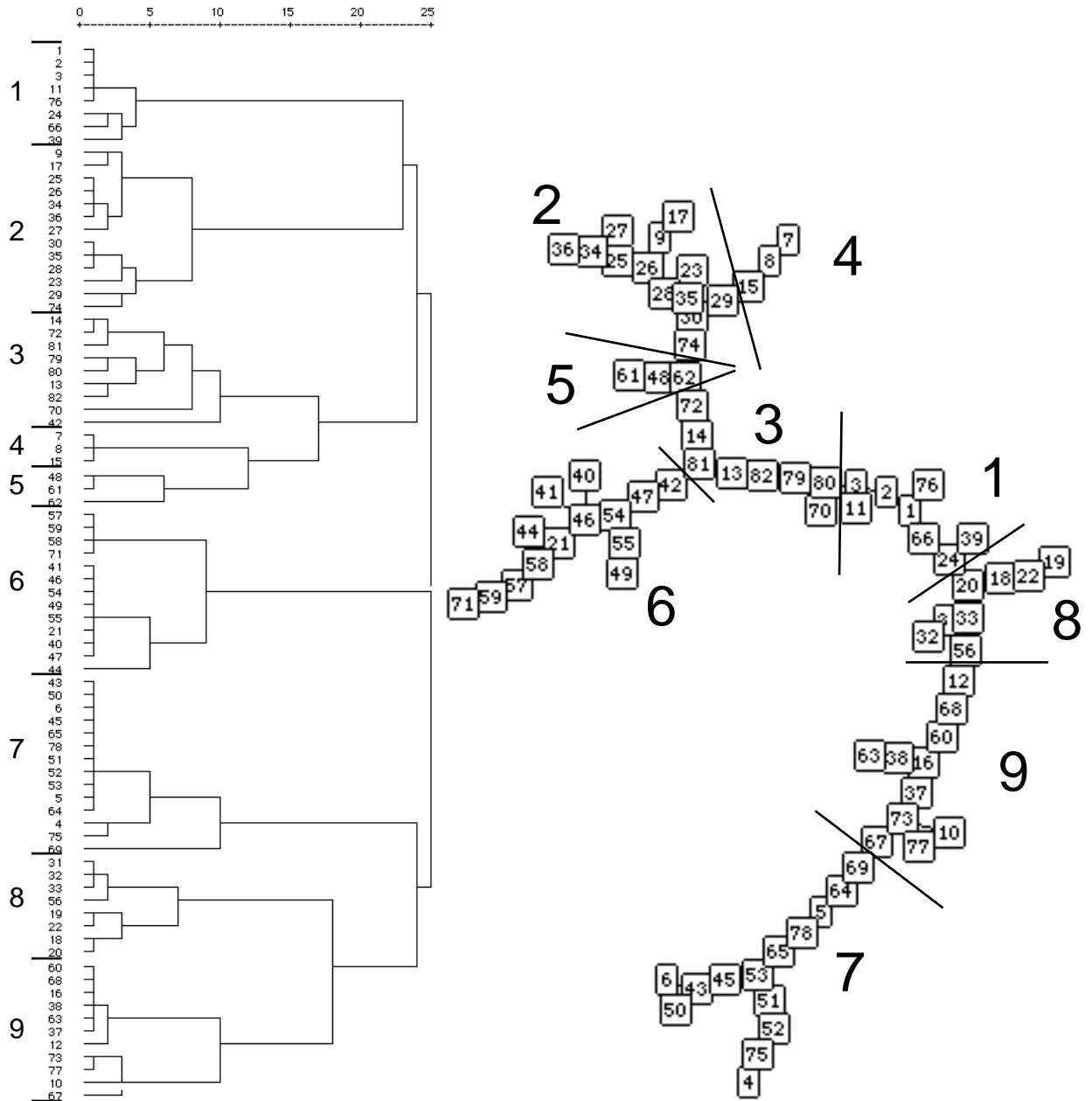
- 1.- Address or telephone of the sports centre.
- 2.-Address or telephone of the Service for Continuous Formation.
- 3.- Address or telephone number of the service for language normalization.
- 4.- Information about Erasmus Grants.
- 5.- Grants for postgraduates.
- 6.-Grants for studying in summer in a foreign country.
- 7.- Services covered by the insurance.
- 8.-Where to call for assistance in case of an accident.
- 9.-Members of the commission for claims about qualifications.
- 10.- Practices in companies.
- 11.-Address of the Foundation University-Enterprise.
- 12.- Courses for foreign visitors.
- 13.- List of university schools..
- 14.- Centres attached to the university.
- 15.- Goal of the insurance.
- 16.- List of postgraduate courses.
- 17.- Normative of postgraduate courses.
- 18.- Enrolment in postgraduate courses.
- 19.- Qualifications for postdoctoral courses.
- 20.- Courses for PhD Students.
- 21.- Certificate of Research Aptitude.
- 22.- Enrolment in PhD Students' courses.
- 23.- Normative about PhD Students' courses.
- 24.- Supervision of the PhD Thesis.
- 25.-Normative regulating the devolution of university fees.
- 26.- Normative about enrolment in the university..
- 27.- Normative for grants: General Chapter.
- 28.-Procedure for continuing studies in other universities.
- 29.-Requirements for direct access in the university when older than 25 years.
- 30.- Steps for enrolment in the university..
- 31.- Program Lingua.
- 32.- Program Tempus.
- 33.- Program Med-Campus.
- 34.- Normative for making the list of courses in an academic year.
- 35.-Normative about the structure and organization of the programs of studies in the university.
- 36.- Normative about the computing services.
- 37.- Computing courses in the first term.
- 38.- Courses of valencian language in the second term.
- 39.- Codes of Centres, Departments and Services in the university.
- 40.- Funding of research in the university.
- 41.- Funding for Research Projects from the local government.
- 42.-Results of the Valencian Centre of Documentation about Narcotics.
- 43.- Grants in foreign countries in engineering.
- 44.-Research Projects and technological development.
- 45.-Funding for projects from the National Government.
- 46.-Funding for the returning of doctors and technologists to research groups.
- 47.-Funding for improvement and mobility of research personnel.
- 48.- Organization of conferences.
- 49.- Submission of proposals to the UE.
- 50.- Grants from the UE..
- 51.- Grants or the program for air transportation studies.
- 52.- Grants for the Foundation Joan Espluges.
- 53.-Funding for Private or Public Institutes of Research.
- 54.- Submission of proposals about research and development.
- 55.- Submission of proposals about scientific and technologic co-operation.
- 56.- Presentation of European Programs.
- 57.- Awards for outstanding research in private institutions.
- 58.- Award for outstanding research from the Royal Academy of Doctors.
- 59.- Award to the best research enterprise.
- 60.-Course on specialization on Clinical Genetics.
- 61.-XXXIV Annual Conference of the Spanish society of Ceramics.
- 62.-Period of Enrolment in the course for pedagogical adaptation.
- 63.- Information about Masters on Commercial Distribution.
- 64.- Money for travelling.
- 65.- Luncheon Vouchers.
- 66.- Telephone of the director of the department of Methodology in Psychology.
- 67.- Work in a foreign country.
- 68.- Masters in the University of Valencia.
- 69.- Discounts for studying English in the British Institute for university students.
- 70.- List of student residences in Valencia.
- 71.- Information about the Graduation Award.
- 72.- List of university courses.
- 73.- Program Comett of internships.
- 74.- Requirement for teaching Spanish in a foreign university and forms to be filled.
- 75.-Results of the Grant for Formation of Research Personnel.

- 76.-Telephone number of the library of the Department of Nursery.
- 77.-Exams for getting a position in the administration.
- 78.- Funding for a conference about AIDS.
- 79.- University Association of mountain-bike.
- 80.- Location of the Botanical Gardens.
- 81.- Date of the Freshman´s day.
- 82.- Theatre groups in the University.

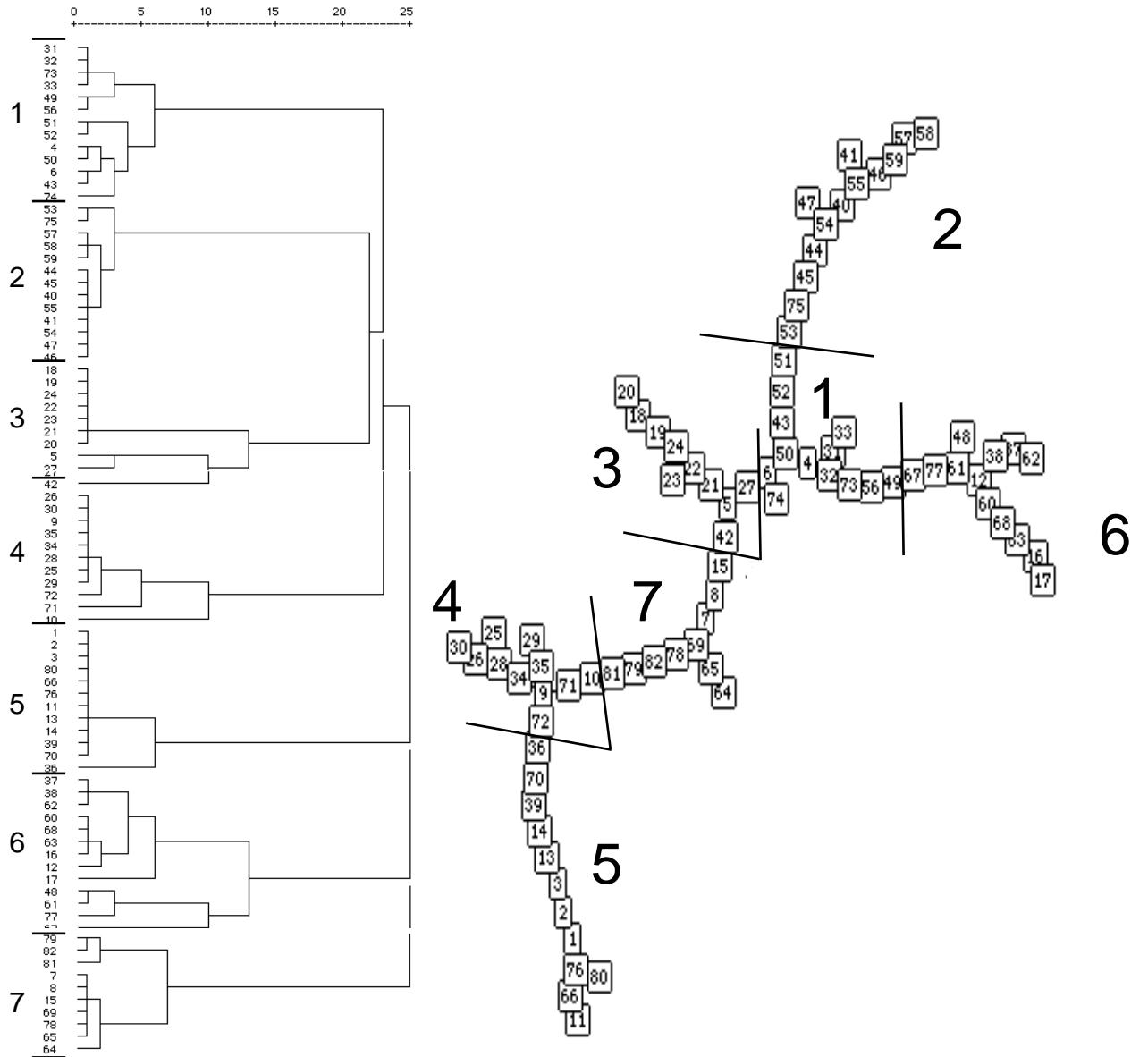
Annex II

Dendrogram and Pathfinder for the groups considered.

Students (ST)



Teaching-Staff Information Services (TS-IS)



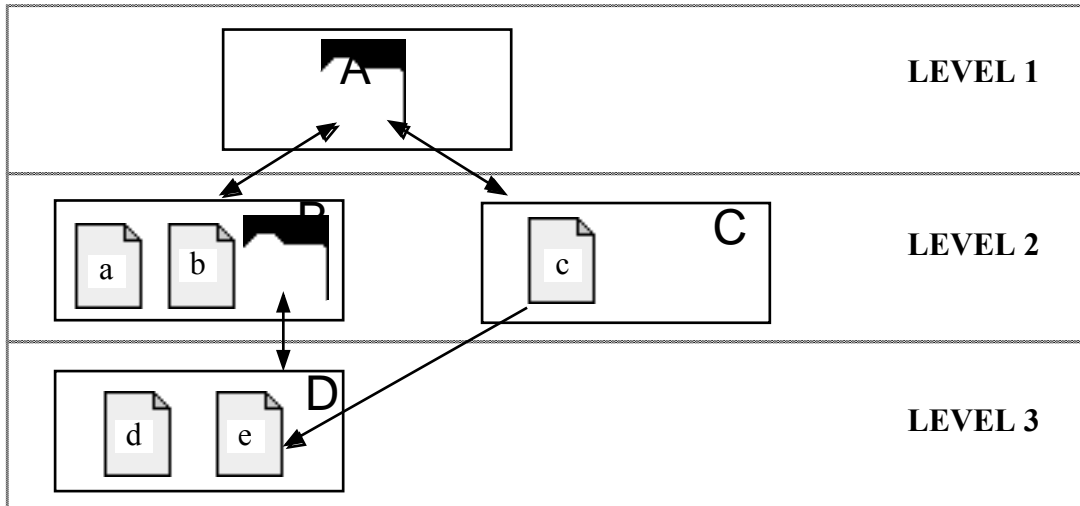
References

- ARABIE, P. 1993, Methodology neither new nor improved, *Contemporary Psychology*, **38**, 66-67.
- ARABIE, P., & HUBERT, L. J. 1992, Combinatorial Data Analysis, *Annual Review of Psychology*, **43**, 169-203.
- BAYLEY, K. D. 1994, *Typologies and Taxonomies: An Introduction to Classification Techniques*, (Sage).
- BENYSH, D. V., KOUBEK, R. J., & CALVEZ, V. 1993, A Comparative Review of Knowledge Structure Measurement Techniques for Interface Design, *International Journal of Human-Computer Interaction*, **5**, 211-237.
- CARROLL, J. D. 1995, "Minimax Length Links" of a dissimilarity matrix and minimum spanning trees, *Psychometrika*, **60**, 3, 371-374.
- COOKE, N. J., & SCHVANEVELDT, R. W. 1988, Effects of computer programming experience in network representations of abstract programming concepts, *International Journal of Man-Machine Studies*, **29**, 407-427.
- DAVISON, M. L. 1983, *Multidimensional Scaling*, (Wiley & Sons).
- EGAN, D. E. 1988, Individual differences in Human-Computer Interaction. In M. Helander (Ed.), *Handbook of Human Computer Interaction*, (North Holland), 543-568.
- CHI, M. T., FELTOVICH, P. J., & GLASER, R. 1981, Categorization and representation of physics problems by experts and novices, *Cognitive Science*, **5**, 121-152.
- GILLAN, D. J., BREEDIN, S. D., & COOKE, N. J. 1992, Network and multidimensional representations of the declarative knowledge of human-computer interface design experts, *International Journal of Man-Machine Studies*, **36**, 587-615.
- HUBERT, L. J. 1985, Combinatorial Data Analysis: Association and Partial Association, *Psychometrika*, **50**, 4, 449-467.
- HUBERT, L. J., GOLLEDGE, R. G., & CONSTANZO, C. M. 1982, Analysis of Variance Procedures Based on a Proximity Measure Between Subjects, *Psychological Bulletin*, **91**, 2, 424-430.
- HUBERT, L. J. 1979, Generalized Concordance, *Psychometrika*, **44**, 2, 135-142.
- HUTCHINSON, J. W. 1989, NETSCAL: a network scaling algorithm for nonsymmetric proximity data, *Psychometrika*, **54**, 25-51.
- KLAUER, K. C. 1989, Ordinal network representation: representing proximities by graphs, *Psychometrika*, **54**, 4, 737-750.
- KOUBEK, R. J. & MOUNTJOY, D. N. 1991, The impact of knowledge representation on cognitive oriented task performance, *International Journal of Human-Computer Interaction*, **3**, 31-48.

- LEGENDRE, P., & FORTIN, M. 1989, Spatial pattern and ecological analysis, *Vegetatio*, **80**, 107-138.
- LEGENDRE, P., & VAUDOR, A. 1991, *The R Package: Multidimensional analysis, spatial analysis*, Département de sciences biologiques, Université de Montreal.
- MCDONALD, J. E., DAYTON, T., & MCDONALD, D. R. 1988, Adapting menu layout to tasks, *International Journal of Man-Machine Studies*, **28**, 417-435.
- ROUNDS, J., TRACEY, T. J., & HUBERT, L. 1992, Methods for evaluating vocational interest structural hypothesis, *Journal of Vocational Behaviour*. 239-259.
- SCHVANEVELDT, R. W. (Ed.), 1990, *Pathfinder Associative Networks: Studies in Knowledge Organization* (Ablex, NJ).
- SCHVANEVELDT, R. W., DEARHOLT, S. W., & DURSO, F. T. 1988, Graph theoretic foundations of pathfinder networks, *Comput. Math. Applic.*, **15**, 4, 337-345.
- VALERO, P. M., CANET, F., & SANMARTÍN, J. 1996, Una comparación entre los métodos Pathfinder y Cluster jerárquico para el análisis de la estructura de la información, *Psicológica*, **17**, 2, 323-336.
- VORA, P. R., HELANDER, M. G. & SHALIN, V. L. 1994. Evaluating the influence of interface styles and multiple access paths in hypertext. In B Adelson, S. Dumais and J. Olson (eds.), *CHI '94*. (Boston, ACM), 323-329.
- YOUNG, F. W., & HAMER, R. M. 1987, *Multidimensional scaling: History, Theory and Applications*, Hillsdale, New Jersey: LEA.

Figures

Figure 1. Distances in a hierarchical system of information organisation.



	a	b	c	d	e
a	0	0	2	1	1
b	0	0	2	1	1
c	2	2	0	3	3
d	1	1	3	0	0
e	1	1	1	0	0

Figure 2. The structure of the matrix of similarities between subjects participating in the study.

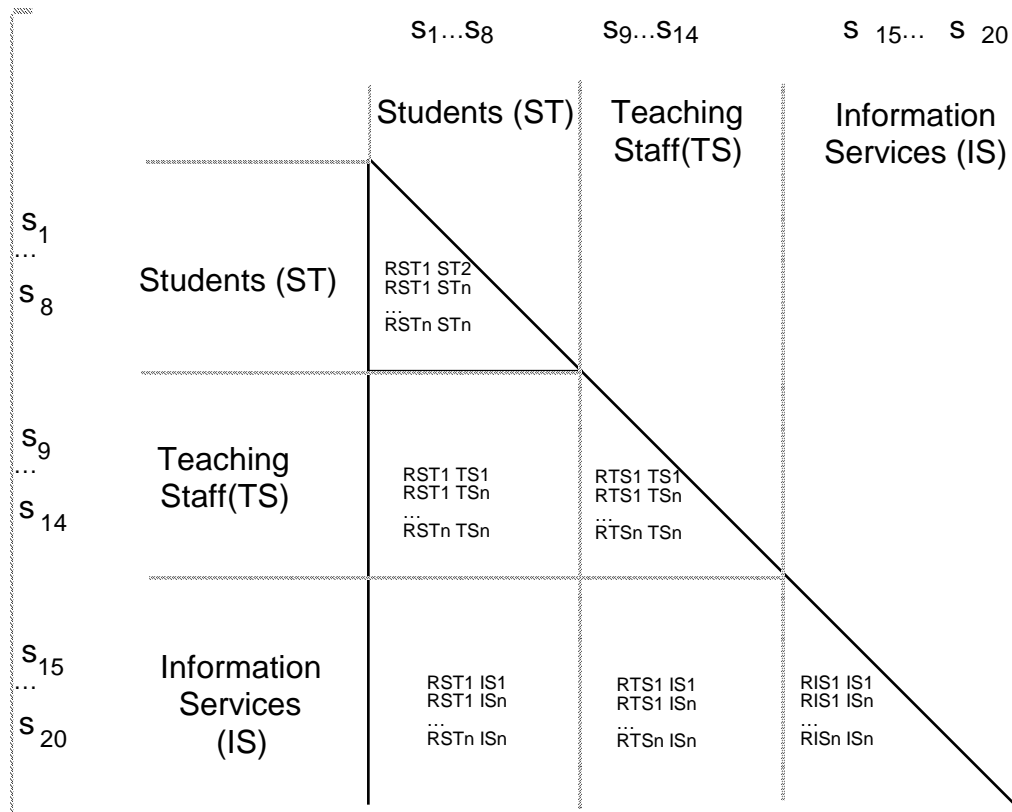


Figure 3. The structure of the matrix of similarities for the comparisons between groups.

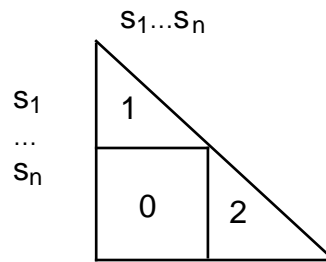


Figure 4: Dendrogram for the cluster analysis on the correlation's between subjects participating in this study. The following symbols are used: • ST; * TS; - IS.

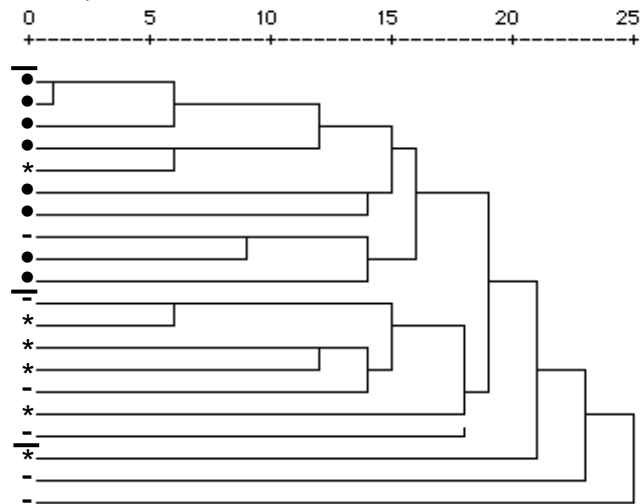


Table 1: Mean correlation's for the groups in question (intragroups) and between the members of the different groups for each one of the comparisons done (intergroups).

	Students	Teaching Staff	Information Services
Students	$\bar{r}_{st\ st} = 0.26$		
Teaching Staff	$r = 0.31273$; $p = 0.00500$ $\bar{r}_{st\ ts} = 0.17$	$\bar{r}_{ts\ ts} = 0.18$	
Information Services	$r = 0.34960$; $p = 0.00300$ $\bar{r}_{st\ is} = 0.14$	$r = 0.06770$; $p = 0.20600$ $\bar{r}_{ts\ is} = 0.14$	$\bar{r}_{is\ is} = 0.12$