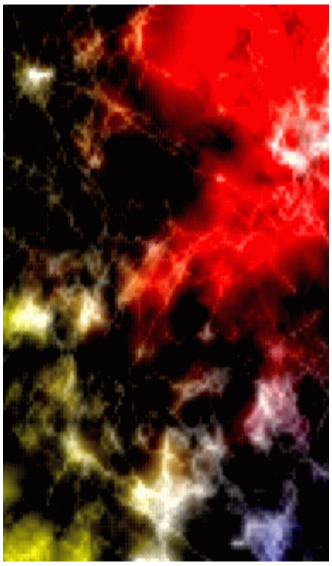
15 October 1995



Cyberspace geography visualization

Mapping the World-Wide Web to help people find their way in cyberspace

Luc Girardin

The Graduate Institute of International Studies, Geneva

<URL:http://heiwww.unige.ch/girardin/cgv/>

Abstract

As cyberspace becomes an integral part of our daily life, its mastering becomes harder. To help, cyberspace can be represented by resources arranged in a multidimensional space. With geographical maps to exhibit the topology of this virtual space, people can have a better visual understanding. In this paper, methods focusing on the construction of lower dimension representations of this space are examined and illustrated with the World-Wide Web. It is expected that this work will contribute to addressing issues of navigation in cyberspace and, especially, avoiding the lost-in-cyberspace syndrome.

Résumé

Alors que le cyberspace envahit notre vie quotidienne, sa maîtrise devient de plus en plus complexe. On peut l'imaginer comme un ensemble de ressources arrangées dans un espace multi-dimensionnel. En utilisant des cartes géographiques pour représente la topologie virtuelle de cet espace, on arrive à mieux le comprendre, le cerner. Dans ce papier, des méthodes se concentrant sur la construction de représentations à dimensions réduites sont étudiées en les appliquant au World-Wide Web. On espère que ce travail contribuera à résoudre les problèmes de navigation dans ce monde virtuel et en particulier à éviter de s'y perdre.

Ubersicht

In einer Zeit, in der der Cyberspace ein integraler Bestandteil unseres täglichen Lebens wird, wird seine Beherrschung zunehmend schwieriger. Zur Erleichterung kann Cyberspace anhand von Quellen, angeordnet in einem multidimensionalen Raum, dargestellt werden. Mit geographischen Karten, die die Topologie dieses künstlichen Raumes aufzeigen, kann das visuelle Verständnis verbessert werden. In dieser Arbeit werden Methoden zur Konstruktion von Darstellungen mit niedriger Dimension dieses Raumes untersucht und anhand des World-Wide Web verdeutlicht. Diese Arbeit trägt somit zur Lösung des Orientierungsproblemen im Cyberspace und insbesondere zur Vermeidung des "Verloren-im-All" Syndrom beim.

Extended abstract

The central goal of this paper is to give information about virtual locations to the actors of cyberspace in order to help them solve orientation issues, i.e. the lost-in-cyberspace syndrome. The approach taken involves low dimensional digital media to create the visualization that can guide you.

The World-Wide Web can be depicted as a graph. Each resource is a vertex and the links are the edges. The distances between pairs of resources is then defined as the shortest path in the graph between them, leading to the creation of a metric. With the ability provided to measure the distances among resources, it becomes possible to represent each resource as a point in a high dimensional space where their relative distances are preserved.

It is clear that a high dimensional space cannot be visualized and thus its dimensionality has to be reduced. To perform this task, the self-organizing maps algorithm is used because it preserves the topological relationships of the original space, conjointly lowering the dimensionality. This creates the ability to map any resources onto a lower dimensional space, while maintaining their order of proximity.

During this non-linear dimensionality reduction, the distances among resources are lost. Since it is primordial that the distances can be evaluated, the unified matrix method is used. By geometrically approximating the vector distribution in the neurons of the self-organizing maps, this method provides a means to analyse the landscape of the mapping of cyberspace.

To permit exploratory analysis of the self-organizing map, the mapping is made onto a two-dimensional visualization media. Note, however, that reduction is also possible, using the proposed method, to a space having an arbitrary dimension. This approach enables the visual display of virtual locations of resources on a landscape, in a fashion similar to geographical maps.

A prototype performing the above task has been developed. Using real information about resources available in the World-Wide Web and their connective structure, various maps have been constructed. Given that the development is in the prototyping stage, it has been possible only to construct maps exhibiting limited numbers of resources. The visualization, comprising some interaction possibilities, is directly made available on the World-Wide Web using forms and sensitive maps, which enable direct retrieval of the resources represented on the maps.

Despite some scalability problems with the current implementation, new developments will soon handle the limitation in information gathering. An implementation model for the construction of the maps on a parallel computer has been proposed. Certainly further improvements are therefore feasible.

The results are encouraging. No major flaw has been detected in the proposed model, and the first users are enthusiasts. It is thus advocated that further research should be done in this direction.

The above mentioned results, including the documentation, are available at

> *<URL:http://heiwww.unige.ch/girardin/cgv/>*

Thousands of accesses to these maps, which show what we would like to call the geography of cyberspace, have already been reported....

Preface

This monograph is a diploma work presented for the Postgraduate Course in Computer science and Telecommunication *(Nachdiplomstudium Informatik und Telekommunication / Formation Postgrade en Informatique et Télécommunication, NDIT/FPIT).*

Chapter 1. "Introduction" explains the concept of cyberspace and discusses the problems of navigability in this virtual world. It also emphizes the usefulness of geographical maps.

Chapter 2. "Problem model" presents a basic model for the cyberspace, the visualization media and the mapping from one to another. The model is explained and formalized mathematically.

Chapter 3. "Solution model" proposes a method based on the self-organizing maps algorithm to transform the elements of cyberspace onto a visualization media, and provides a method to visualize the landscape of the map.

Chapter 4. "Results" presents various maps that have been constructed based on real data collected in the World-Wide Web. Basic information on the construction of the prototype is also given.

Chapter 5. "Possible enhancements" discusses possibilities for improving the actual model and its implementation. In particular, a model of scalability based on parallel computing is proposed.

Chapter 6. "Conclusion" synthesizes the work and draws overall conclusions.

To help readers with the definition of some terms used in this paper, a glossary, beginning on page 37, is provided.

To increase readability, only essential references are provided in the text. For further study, readers should refer to the annotated bibliography beginning on page 41.

I'm indebted to a large number of people who have helped greatly in completing this research. I'm very thankful for the support of Lorenz Müller and Jean-Gabriel Gander, the two supervisors, and Boi Faltings, the expert in this work. I am grateful for their interesting discussions, corrections, and suggestions to Samantha Anderson, René Bach, Nicolas Droux, Claude Fuhrer, Catherine Kuchta, Daniel Liebhart, Jennifer Milliken, Hervé Sanglard, Patricia Weitsman and Andrew Wood. Special thanks are due to my colleagues Marielle Schneider, Edgardo Amato and Wilfred Gander.

This document has been written using FrameMaker. The bibliographic references are made according to the International Standard Bibliographic Description.

Table of contents

1. Introduction

The World-Wide Web [Hughes, 1994] is actually the incarnation of the concept of cyberspace[1][Gibson, 1984][Benedikt, 1991], a theatre of complex interactions. Cyberspace can be seen as the latest stage in the evolution of Popper's World 3 [Popper, 1979], the world of objective, real, and public structures. The World-Wide Web project, an Internet-based hypermedia initiative for global information sharing, has been inaugurated at CERN (European Laboratory for Particle Physics) in 1989 by Tim-Berner Lee and has become increasingly popular.

As we can move in our real world, we can wander in cyberspace. In this virtual world, people are able to navigate through "a common mental geography", "a nowhere space" or "a consensual hallucination". This space is multidimensional and therefore seems considerably different from the notions of physical space that many of us have. This multidimensionality makes it very difficult to determine the overall structure of the World-Wide Web. Since information about the orientation is globally poor, the so-called lost-in-cyberspace syndrome has become an important problem, limiting the cyberspace navigability.

Through its maps and more recently satellite imaging, geography,[2] has played a major role in the analysis of human activity. People are easily able to explore cities and navigate through countries they have never visited before thanks to geographic tools. Although the physical earth has been completely mapped, no such maps exist for cyberspace.

Visualization[3] creates the possibility of communicating large amounts of information to the human visual system. If information about an emergent topology of the World-Wide Web can be found, an approximate representation can be built in a dimension appropriate for visualization.

The project presented here describes how the geographical features of cyberspace can be extracted and visualized. To this end, the following chapters develop a mapping of the World-Wide Web in a medium of low dimension in order to help people have visual information about virtual locations.

1. William Gibson coined the term cyberspace when he sought a name to describe his vision of a global computer network, linking all people, machines, and sources of information in the world through which one could navigate as through a virtual space. The original definition from his futuristic novel *Neuromancer* is:

   "Cyberspace. A consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts... A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data. Like city lights receding...." [Gibson, 1984]

2. The general definition of geography is the topographical features of any complex entity.
3. Visualization is the process of transforming information into a visual form, enabling users to observe the information.

2. Problem model

2.1 Cyberspace representation

The World-Wide Web is currently the most popular incarnation of the concept of cyberspace. It makes use of Uniform Resource Locators (URLs) [Connoly, 1995b] to identify resources, most often documents. The operation of the World-Wide Web relies mainly on hypermedia structures as a means of navigation for users. This is done by anchoring links to other resources through the use of the HyperText Markup Language (HTML) [Connoly, 1995a]. Thus, any resource can be linked by reference to any other.

2.1.1 Representation with a graph

Therefore, we can see cyberspace as a finite set $A = \{a_1, a_i, \ldots, a_n\}$ of $|A|$ resources with a relation $\alpha \subseteq A \times A$ between pairs $(a_i, a_j) \in \alpha$ of linked resources. It is possible to model this system with a connected graph $G = (A, \alpha)$ where $A = A(G)$ represent the vertices (nodes) and $\alpha = \alpha(G)$ the edges (undirected arcs) between vertices of the graph. The size of G is the number of edges, thus $n = |G| = |A(G)| = |A|$.

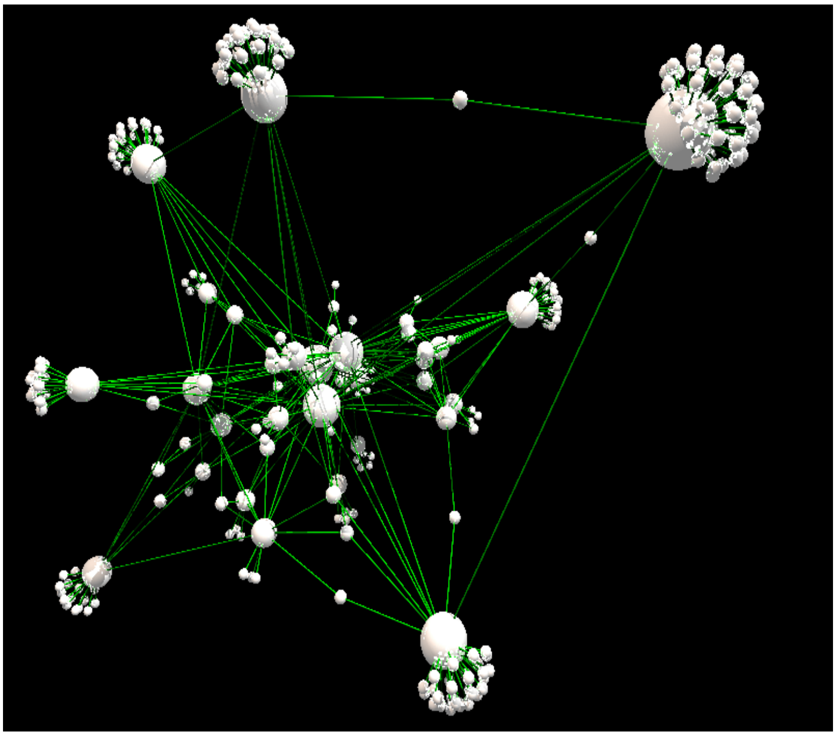
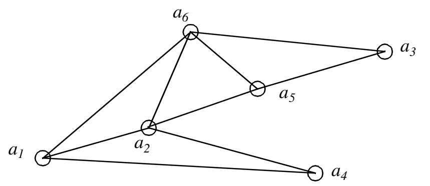


FIGURE 1. **Representation example of cyberspace as a graph [Wood et al., 1995].**

The information in the graph G may also be expressed in a variety of ways in matrix form. There is one such matrix, the adjacency matrix, that is especially useful. An adjacency matrix $S = S(G)$ of the graph G is of size $n \times n$. The entries in the adjacency matrix, s_{ij}, records which pairs of nodes are adjacent. If nodes a_i and a_j are adjacent, then $s_{ij} = 1$, and if nodes a_i and a_j are not adjacent, then $s_{ij} = 0$. The entries on the diagonal, values of s_{ii}, are undefined, because we do not allow loops in the graph.



$$S = \begin{bmatrix} - & 1 & 0 & 1 & 0 & 1 \\ 1 & - & 0 & 1 & 1 & 1 \\ 0 & 0 & - & 0 & 1 & 1 \\ 1 & 1 & 0 & - & 0 & 0 \\ 0 & 1 & 1 & 0 & - & 1 \\ 1 & 1 & 1 & 0 & 1 & - \end{bmatrix}$$

FIGURE 2. A graph G and its adjacency matrix.

The following elements are introduced to extract features and components of the graph G:

- the set of adjacent nodes of the vertex a:

  $\Gamma: (G, A) \to \wp(A)$, $\qquad (G, a_i) \to \Gamma(G, a_i) = \Gamma^G_{a_i}$

  where $\wp(A)$ denotes the set of the parts of A;

- the degree of a vertex a, which is the number of edges incident to the vertex. In the adjacency matrix the nodal degrees are equal to either the row sums or the column sums. This degree can be seen, for our purposes, as a characteristic of a resource.

  $$\gamma: (G, A) \to \aleph, \qquad (G, a_i) \to \gamma(G, a_i) = \left|\Gamma^G_{a_i}\right| = \sum_{j=1}^{n} s_{ij} = \gamma^G_{a_i}.$$

2.1.2 Spatial representation

The dissimilarity between two resources can be defined by the length of the shortest edge-sequence (path) between them:

$$d: A \times A \to \aleph, \qquad (a_i, a_j) \to d(a_i, a_j) = d_{ij}$$

with

$$d_{ii} = 0,$$

$$d_{ij} \geq 0,$$

$$d_{ij} = d_{ji},$$

$$d_{ik} \leq d_{ij} + d_{jk}, \qquad \forall a_i, a_j, a_k \in A$$

and

$$d_{ij} = 1, \qquad \forall (a_i, a_j) \in \alpha, a_i \neq a_j.$$

This is in fact the geodesic distance and it can be calculated by building a power matrix, starting with $p = 1$. When $p = 1$, the power matrix is the adjacency matrix, so that if $s^{[1]}_{ij} = 1$, the resources are adjacent, and the distance between them equals 1. If $s_{ij} = 0$ and $s_{ij}^{[2]} > 0$, then the shortest path is of length 2 and so forth. Consequently, the first power p for which the s_{ij} is non-zero gives the length of the edge-sequence and is equal to d_{ij}. Mathematically,

$$d_{ij} = \min_p (s_{ij}^{[p]} > 0) .$$

Note that $s_{ij}^{[n]}$ is the number of paths between the resources a_i and a_j.

With such a metric defined, it is possible to construct a distance (dissimilarity) matrix $D(G) = D = [d_{ij}]$ of the graph G, composed of vectors $d_i = (d_{i1}, d_{ij}, \ldots, d_{in})$ of n dimensions. Therefore, each vector d_i gives an unique representation of each resource as a point in an n-dimensional space.

$$D = \begin{bmatrix} 0 & 1 & 2 & 1 & 2 & 1 \\ 1 & 0 & 2 & 1 & 1 & 1 \\ 2 & 2 & 0 & 3 & 1 & 1 \\ 1 & 1 & 3 & 0 & 2 & 2 \\ 2 & 1 & 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 2 & 1 & 0 \end{bmatrix} \qquad d_4 = (1, 1, 3, 0, 2, 2)$$

FIGURE 3. Example of a distance matrix

2.2 Visualization media representation

In current digital systems, visualization is usually made over two-dimensional media. Pictures are composed of patterns of pixels (picture elements). Although the configuration of the pixels can be constructed in an arbitrary fashion, they are normally represented as lattices of squared cells.

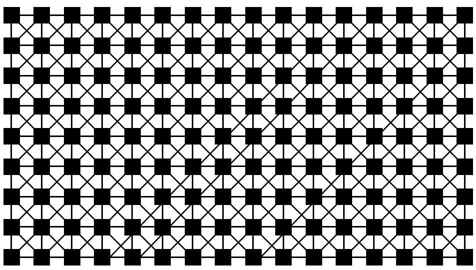


FIGURE 4. Representation example of a visualization media as a graph.

By considering the neighborhood of each pixel, modelling a connected pattern with a graph is straightforward. Calling this graph H and using the definitions presented earlier, we can introduce the following elements:

- the graph made of the set of pixels B and their connections β : $H = (B, \beta)$,
- the number of elements in the graph H : $m = |H| = |B(H)|$,
- the adjacency matrix of the graph H : $T = T(H) = [t_{rs}]$,
- the set of adjacent pixels of a pixel b_r : Γ_r^H ,
- the degree of a pixel b_r : γ_r^H ,
- the distance between two pixels b_r and b_s : $\delta_{rs} = \min_p t_{rs}^{[p]} > 0$
- the distance matrix: $D(H) = \Delta = [\delta_{rs}]$.

As an extra definition, we introduce the characteristic κ_r of a pixel r :

$$\kappa : B \to \mathrm{N}, \qquad b_r \to \kappa(b_r) = \kappa_r .$$

2.3 Mapping cyberspace over a visualization media

By mapping the graph G over the graph H, a representation of cyberspace over a low-dimensional media becomes possible. Two approaches can be followed. The first is to find directly a mapping of the graph G over the graph H by using the local information known by each resource, i.e. the adjacency matrix. The second is to give a spatial representation of the graph G so as to map its points in a space of a lower dimension. This mapping leads to the possibility of finding by use of a distance matrix, corresponding nodes in the graph H.

Since we want to represent the topographical features, it is important to consider a proper scheme for representing the characteristics of the resources, defined in the graph G, in the graph H.

2.3.1 Mapping

To accomplish visualization, a representation over low dimensional graphical media is to be done. The goal is therefore to find a proper representation of cyberspace by conserving its topological features on a media composed of discrete elements.

Suppose a mapping can be defined by an equivalency class between A and B:

$$\Phi : A \to A/\sim = B, \qquad a_i \to \Phi(a_i) = \Phi_i = [a_i] = b_r$$

we can talk about the pre-image:

$$\Phi^{-1} : B \to A, \qquad b_r \to \Phi^{-1}(b_r) = \Phi_r^{-1} = A_i \subseteq A \ | \ \Phi(a_i \in A_i) = b_r .$$

.

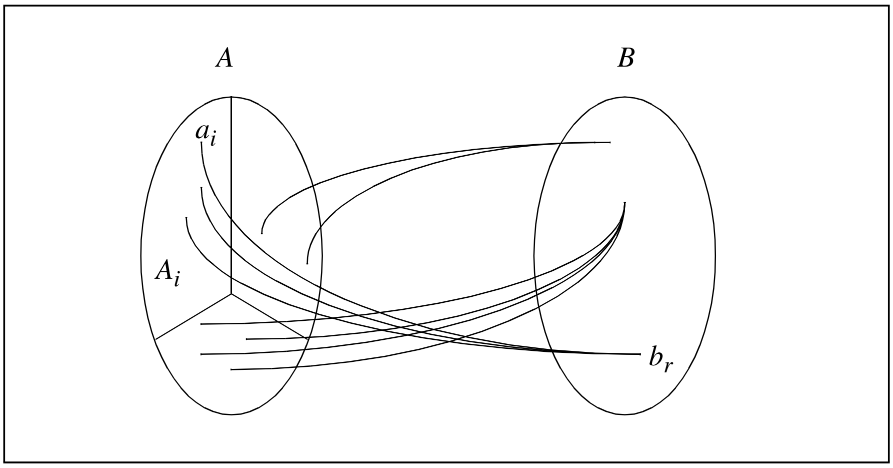


FIGURE 5. **Mapping between A and B**

2.3.2 Morphisms

To give a proper view of the structure of cyberspace over a medium, various requirements for the mapping are described below, in term of morphisms and from weak to strong constraints.

An exact match from the set A to set B is a mapping that has a corresponding relation tuple (element) in β for each relation in α. A mapping fulfilling this requirement is called a homomorphism from α to β such that

$$(\Phi \bullet \alpha) \subseteq \beta, \qquad \forall (a_i, a_j) \in \alpha \Rightarrow (\Phi_i, \Phi_j) \in \beta .$$

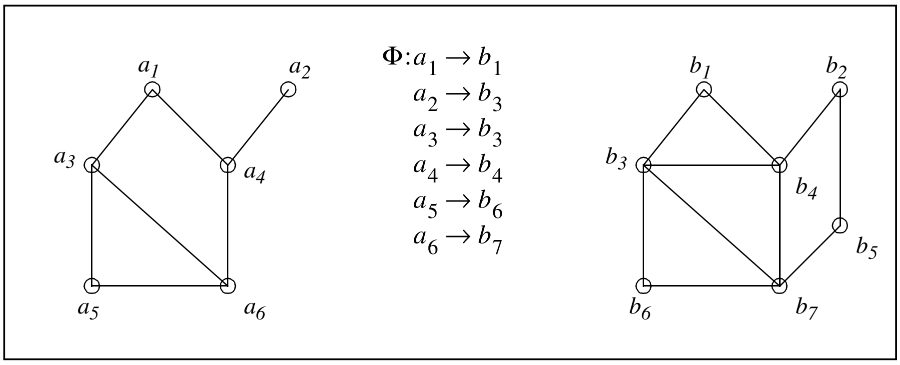


FIGURE 6. **A homomorphism**

Consider that a mapping exists wherein there is a one-to-one correspondence between the vertices in G and the vertices in a subgraph of H such that a pair of vertices are adjacent in G if and only if the corresponding pair of vertices are adjacent in the subgraph of H. This is in fact the condition for a monomorphic mapping:

$$(\Phi \bullet \alpha) \subseteq \beta \text{ and } \Phi \text{ is 1-1}, \qquad \forall (a_i, a_j) \in \alpha \Leftrightarrow (\Phi_i, \Phi_j) \in \beta$$

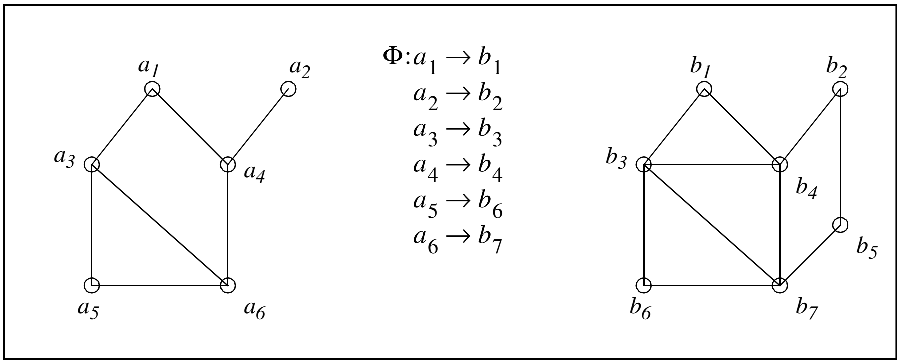


FIGURE 7. A monomorphism

which can also be described as a isomorphic mapping to a subset β' from the relation β such that

$$(\Phi \bullet \alpha) \subseteq \beta' \text{ and } (\Phi^{-1} \bullet \beta') \subseteq \alpha .$$

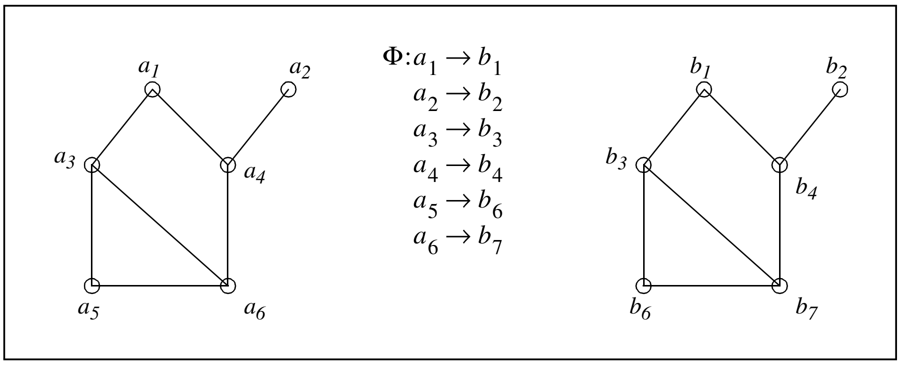


FIGURE 8. An isomorphism

Note that all these morphism problems have been shown to be NP-complete and therefore cannot be solved exactly in polynomial time.

2.3.3 Metrics

A contrasting approach is to define a requirement based on the distance among elements. This approach clearly gives more freedom, but also increases the computational complexity since the transformation will now be based on the distance matrix.

Suppose that the goal of the transformation has a central feature of obtaining a monotone relationship between distances. Then only the rank order of the dissimilarities has to be preserved by the transformation. Hence, the metric is abandoned during the mapping. Therefore the transformation must obey the monotonicity constraint

$$d_{ij} \le d_{kl} \Leftrightarrow \delta(\Phi_i, \Phi_j) \le \delta(\Phi_k, \Phi_l)\,, \qquad \forall a_i, a_j, a_k, a_l \in A\,.$$

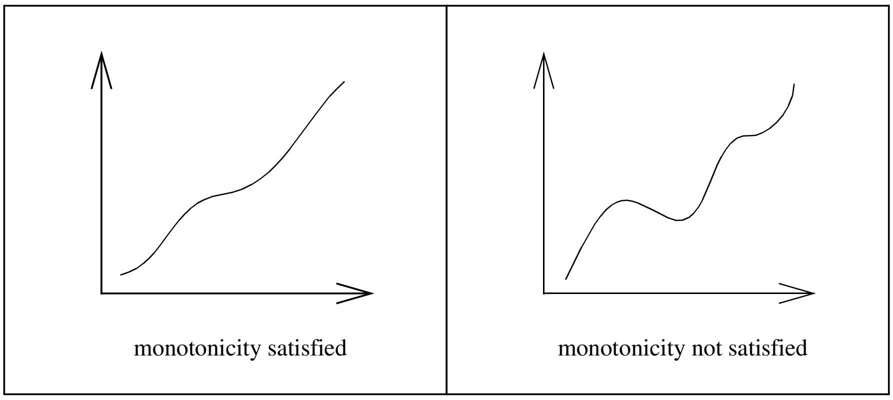


FIGURE 9. Monotonicity constraint

If the metric nature of the transformation is to be preserved, a configuration will have to satisfy

$$d_{ij} = f(\delta(\Phi_i, \Phi_j))\,, \qquad \forall a_i, a_j \in A$$

where f is a monotonic function of the distance; a possible straighforward example could be

$$f(\delta(\Phi_i, \Phi_j)) = E\delta(\Phi_i, \Phi_j) + \varepsilon\,.$$

The stronger constraint that we can put on the mapping is the isometry, having then a perfect preservation of the topology. An isometry is defined by

$$\delta(\Phi_i, \Phi_j) = d_{ij}, \qquad \forall a_i, a_j \in A\,.$$

It is obvious that a mapping that satisfies the isometry also conserves the monotonicity among distances.

2.3.4 Characteristics mapping

With defined mapping, the features or characteristics of any resources can be calculated for a pixel b_r. For example, this can be the sum of the degrees of the corresponding resources:

$$\kappa_r = \sum_{a_i \in \Phi_r^{-1}} \gamma^G{}_{a_i}\,.$$

3. Solution model

3.1 Information gathering

The World-Wide Web, is currently organized in a client-server fashion. Therefore, information gathering has to be undertaken over the network. A technical description of the actual organization is explained in annex A. “Information gathering and the World-Wide Web”.

3.1.1 Exploration strategy

In section 2.1, cyberspace was modelled as a graph. Two classical graph-traversal algorithms are appropriate for gathering information about the connective structure of this graph: the depth-first search and the breadth-first search.[1] Since they are made with adjacency lists, both searches require time proportional to $|A| + |\alpha|$. The graph can be represented with an adjacency-structure (linked-lists).

3.1.2 Adjacency matrix construction

To construct the adjacency matrix, the resources are identified by using integers between 1 and $|A|$. The construction is straightforward and demands $|\alpha|$ steps, resulting in a matrix of $|A|^2$ bits.

3.1.3 Distance matrix construction

From the adjacency matrix, the shortest edge-sequences to and from every resources are calculated. This is done by applying the classical shortest-path algorithm $|A|$ times, resulting in a complexity of $O((|\alpha| + |A|)|A|\log|A|)$. Another possibility is to use Floyd’s algorithm which solves the all-pairs shortest-path problem in $O(|A|^3)$ [Sedgewick, 1989].

3.2 Mapping

Having gathered the data, we need to present them on a media suitable for visualization. For this, a mapping has been defined in section 2.3.1. Because actual digital systems use an array of squared pixels, the presentation below focuses on this kind of organization to simplify comprehension for the reader. Note, though, that a mapping to any kind of organization can be extrapolated.

A naive approach to this problem would be to find a configuration in H that preserves the connective structure of the graph G (see section 2.3.2 “Morphisms” on page 15). Depending on the data set, such a mapping can only give poor results, since elements of cyberspace do not have, by far, the same connectivity as elements of visualization media.

As specified in section 2.1.2, each resource is represented as a point in data space $\Re^n$ where n is the number of elements of G. Thus, a constraint to

1. Other exploration strategies are presented in section 5.1 “Improved search strategy” on page 33.

dimensionality reduction can be established (see section 2.3.3 “Metrics” on page 16).

3.2.1 Metric multidimensional scaling

The attempt to find a configuration where the metric nature of the transformation is conserved (see section 2.3.3 “Metrics” on page 16) is called metric multidimensional scaling [Cox et al., 1994] or linear dimensionality reduction. Various methods to complete this task have existed for a long time . The most famous technique is principal components (coordinates) analysis (Karhunen-Loève expansion) [Jolliffe, 1986][Auray et al., 1990a]. Other techniques include least squares scaling and Critchley’s method [Cox et al., 1994].

It is obvious that conserving the metric nature of the distances is important. However, since it is implausible that a linear relationship can be found during the transformation, the above mentioned methods do not perform well on high-dimensional dimensionality reduction. Thus, alternatives to metric multidimensional scaling have to be investigated.

3.2.2 Non-metric multidimensional scaling

If the metric nature of transformation is abandoned, non-metric multidimensional scaling [Kruskal and Wish, 1981][Wilkinson et al., 1992][Cox et al., 1994] or non-linear dimensionality reduction [Li et al., 1995] can be defined as a mapping which only has to obey the monotonicity constraint. Popular methods for doing this transformation, which differ mainly on the cost function[2] to be minimized, are Kruskal’s approach [Kruskal and Wish, 1981], the Guttman approach [Cox et al., 1994] and Sammon’s mapping (non-linear mapping) [Kohonen, 1995][Li et al., 1995].

Although these methods could lead to a potential solution to our mapping problem[3], a fairly new method, the self-organizing map algorithm, has proved to outperform them [Li et al., 1995].

3.2.3 Self-organizing maps

Self-organizing maps [Kohonen, 1995][Müller and Wyler, 1994][Blayo, 1995] are inspired from biology. They are designed to behave, for example, like the somatotopic map of the motor nerves and the tonotopic map of the auditory region. The self-organizing map algorithm, first introduced by Kohonen, is an unsupervised (self-organizing) neural network composed of an input layer and a competitive neural layer. For our goal, the most interesting property of this network is that the feature map preserves the topology of stimuli according to their similarity.

2. The cost function tries to estimate how well a configuration satisfies the requirements. It is a prerequisite for creating an optimization model

3. Keeping in mind that abandoning the metric nature of the transformation will not be satisfactory for our particular problem, and thus a way to recover it will have to be sought.

We can use self-organizing maps to lower the dimensionality and preserve the topological features of the data. To complete this task, we present to the input layer the distance vectors, i.e. the coordinates of each resource as a point in an n-dimensional space,

$$d_i = (d_{i1}, d_{i2}, \ldots, d_{in}).$$

The neurons in the competitive layer are in fact the pixels b_r and a weight (reference) vector w_r is associated with them:

$$w_r = (\omega_{r1}, \omega_{r2}, \ldots, \omega_{rn}) \in \Re^n$$

with $r = 1, 2, \ldots, m$.

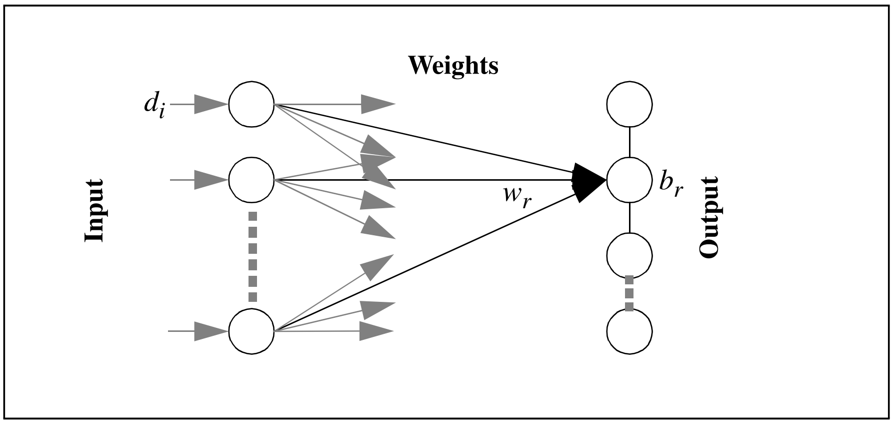


FIGURE 10. Structure of the self-organizing maps neural network (one-dimensional case)

The self-organizing map algorithm in the learning process stage[4] can be summarized as follow:

1. Initialization of the reference vectors;
2. Presentation of a vector to the input layer;
3. Detection of the neuron having the closest reference vector;
4. Modification of the reference vectors of the neurons surrounding the winner;
5. Repetition of steps 2-5 until the number of required iterations has been reached.

According to the Euclidean distance[5] between an input vector d_i and a weight vector w_r of the neuron b_r:

$$\|d_i - w_r\| = \sqrt{\sum_{j=1}^{n} (d_{ij} - \omega_{rj})^2}$$

4. The learning process usually consists of two stages: a coarse-adjustment pass and a fine-adjustment pass.

a winning neuron b_k is defined as the one which has the closest weight vector:

$$\|d_i - w_k\| = min_r(\|d_i - w_r\|)$$

During the learning period, a neighborhood function $C(k)$ is defined to activate neurons that are topographically close to the winning neuron. This function is usually

$$C(k)^{(t)} = C(k)^{(0)}\left(1 - \frac{t}{T}\right)$$

where $C(k)^{(0)}$ is the initial neighborhood (radius)[6].

The reference vectors of the neurons surrounding the winner are modified as follows:

$$w_r^{(t+1)} = \begin{cases} w_r^{(t)} + \eta^{(t)}(d_i^{(t)} - w_r^{(t)}) & \text{if } b_r \notin C(k) \\ w_r^{(t)} & \text{if } b_r \in C(k) \end{cases}$$

with $\eta^{(t)}$ a monotonically decreasing function, for example:

$$\eta^{(t)} = \eta^{(0)}\left(1 - \frac{t}{T}\right)$$

$\eta^{(0)}$ being the initial learning rate[7] and T the number of training operations.[8]

Note that self-organizing maps are performed in a way similar to the k-means [Belaïd and Belaïd, 1992] algorithm used in statistics. Although, the latter has been shown to perform differently and less satisfactory than the first [Ultsch, 1995]

It must be noted that no proof of convergence of the self-organizing maps algorithm, except for the one-dimensional case, has yet been presented [Kohonen, 1995]. Although, it is important to evaluate the complexity of the algorithm. Since the convergence has not be formally proved, we must rely on empirical experiments to determine t. Thus, for $t = 500m$, it has been

5. Note that to respect our model, the so-called city-block (Manhattan) distance should be used. Although, the Euclidean distances gives better visual effects in a two-dimensional space. Therefore, to fully respect our model, the distance should be defined by

$$\|d_i - w_r\| = \sum_{j=1}^{n} |d_{ij} - \omega_{rj}| .$$

6. This is usually about half the diameter of the network during the coarse-adjustment pass and substantially during the fine-adjustment pass.
7. Typically 0.5 during the coarse-adjustment pass and 0.02 during the fine-adjustment pass.
8. For statistical accuracy, T should be at least $500m$ during the final convergence phase.

shown that the complexity of the algorithm is of the order $O(m^2 n)$ [Wyler, 1994].

Having completed the learning process, the mapping can be processed by calculating for each input distance vector the winning neuron. We have then a mapping function which for each resource a_i returns a pixel b_r:

$$\Phi_i = \operatorname{argmin}_r (\|d_i - w_r\|) .$$

3.3 Landscape representation

To enable visualization, a topologically-organized map is not sufficient. The problem resides, since only order is conserved in the fact that the distances among elements are lost. Thus, a method to visualize the structure of the self-organizing map is followed. This can be done by giving an approximation of the weight vector distribution in the self-organizing map. Two methods can be used to complete this task: the unified matrix and the s-diagram [Siemon and Ultsch, 1992]. Since the latter gives a poor approximation of the weight vector distribution, only the first one will be examined.

3.3.1 Unified matrix method

Consider the self-organizing map to be a two-dimensional array with a rectangular lattice topology. Let the matrix of neurons, of size $X \times Y$, be denoted $\left[b_{x,y}\right]$ and the matrix of weights be denoted $\left[w_{i_{x,y}}\right]$. We can now define the three following distances:

$$dx(x,y) = \|b_{x,y} - b_{x+1,y}\| = \sqrt{\sum_i (w_{i_{x,y}} - w_{i_{x+1,y}})^2} ;$$

$$dy(x,y) = \|b_{x,y} - b_{x,y+1}\| = \sqrt{\sum_i (w_{i_{x,y}} - w_{i_{x,y+1}})^2} ;$$

$$dxy(x,y) = \frac{\dfrac{\|b_{x,y} - b_{x+1,y+1}\|}{\sqrt{2}} + \dfrac{\|b_{x,y+1} - b_{x+1,y}\|}{\sqrt{2}}}{2} = \frac{\dfrac{\sqrt{\sum_i (w_{i_{x,y}} - w_{i_{x+1,y+1}})^2}}{\sqrt{2}} + \dfrac{\sqrt{\sum_i (w_{i_{x,y+1}} - w_{i_{x+1,y}})^2}}{\sqrt{2}}}{2} .$$

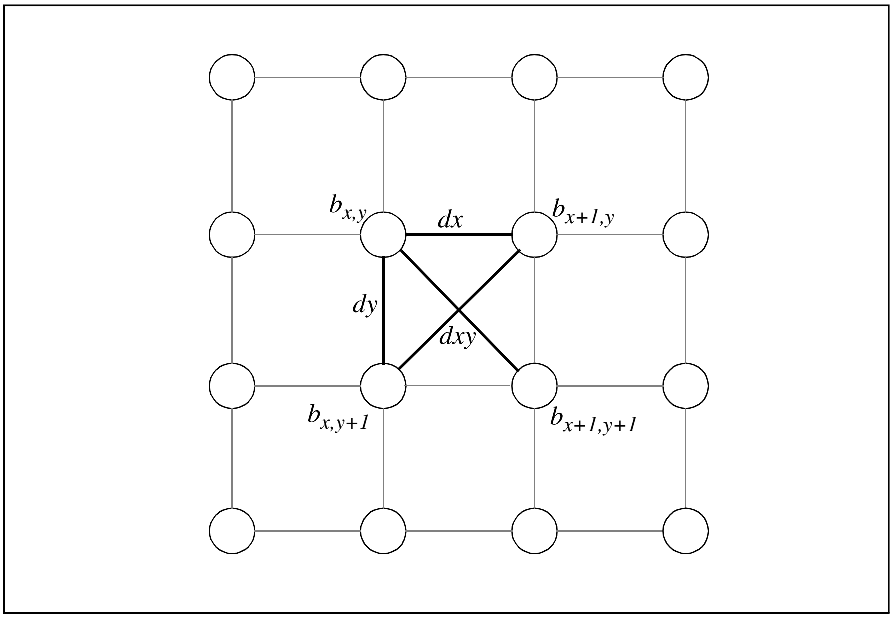


FIGURE 11. The different distances

The so-called unified matrix method [Ultsch and Siemon, 1989][Ultsch, 1993] has been proposed to combine the three distances into one matrix $U = \left[u_{x,y}\right]$ of size $2X - 1 \times 2Y - 1$. For the column positions $2x - 1$, $2x$ and $2x + 1$, and for row positions $2y - 1$, $2y$ and $2y + 1$, the components of the matrix take their value as follow:

$$U = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & dxy(x-1, y-1) & dy(x-1, y) & dxy(x-1, y) & \dots \\ \dots & dx(x, y-1) & du(x, y) & dx(x, y) & \dots \\ \dots & dxy(x, y-1) & dy(x, y) & dxy(x, y) & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Mathematically, this gives:

$$u_{2x+1, 2y} = dx(x, y) ;$$

$$u_{2x, 2y+1} = dy(x, y) ;$$

$$u_{2x+1, 2y+1} = dx(x, y) ;$$

$$u_{2x, 2y} = du(x, y) .$$

with $du(x, y)$ being the median[9] of the surrounding elements, thus

$$du(x, y) = \tilde{x} = \begin{cases} x_{(l+1)/2} & \text{if } l \text{ odd} \\ \dfrac{x_{l/2} + x_{(l+1)/2}}{2} & \text{if } l \text{ even} \end{cases}$$

9. This is obviously not the only possibility. The use of the mean value could also make sense.

where $x_1, x_2, \ldots, x_l$ denotes the surrounding elements arranged in increasing order of magnitude.

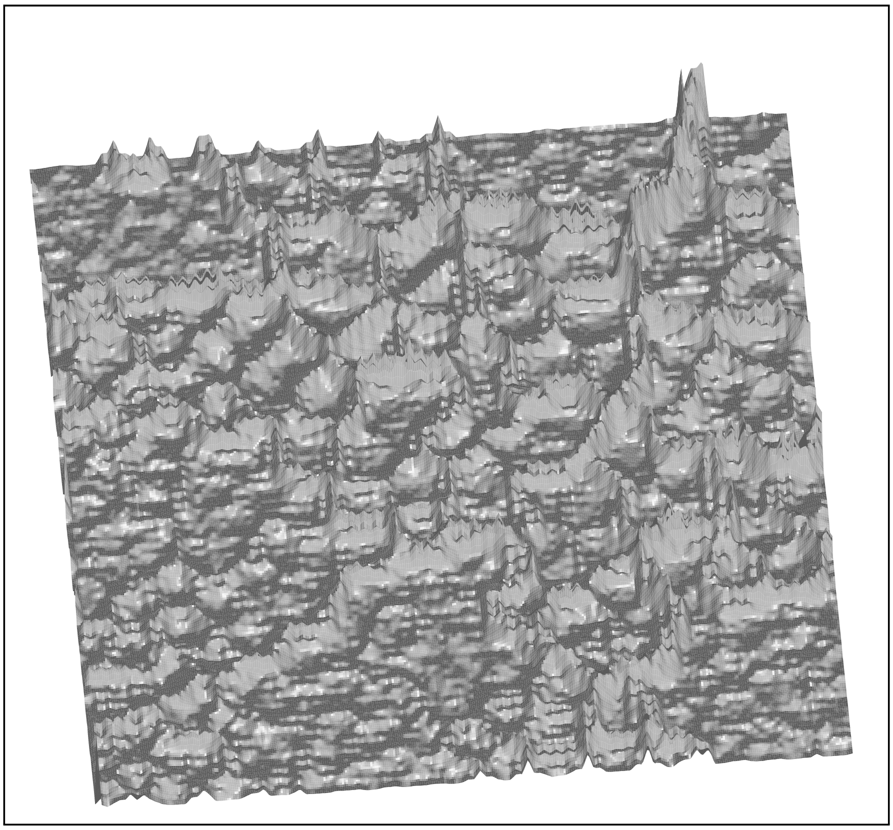


FIGURE 12. Representation of the landscape in 2.5 dimensions.

3.3.2 Visualization

Since a model for the representation of the landscape and of the resources has been established, we can now put into final form by giving a simple scheme for their visualization on a color medium.

For this purpose, we can model the visualization medium, for a RGB color model,[10] as composed of three planes. Each plane corresponds to one color channel, i.e. the red, green and blue channels, which control the intensity of each color. Then, for a 2.5-dimensional[11] representation, we can assign, for example, to a visualization media

$$RGB_{xy} = ([0-255], [0-255], [0-255])_{xy}$$

the following values:

$$RGB_{xy} = (r_{xy}, g_{xy}, b_{xy})$$

10.The RGB model is used here for simplicity. Other color models, like HSV (Hue, Saturation, and Value), resemble more closely the real color system and should be used instead.

11.The color represent here the additional half dimension.

with r_{xy}, g_{xy}, and b_{xy} being

$$r_{xy} = \begin{cases} 255 \cdot \left(\dfrac{\kappa^{(R)}_{x/2,\, y/2} - \min_{\kappa^{(R)}}}{\max_{\kappa^{(R)}} - \min_{\kappa^{(R)}}} \right) & \text{if } \Phi^{-1}_{x/2,\, y/2} \neq \{\varnothing\} \\ 255 \cdot \left(\dfrac{u_{xy} - \min_u}{\max_u - \min_u} \right) & \text{if } \Phi^{-1}_{x/2,\, y/2} = \{\varnothing\} \end{cases},$$

$$g_{xy} = \begin{cases} 255 \cdot \left(\dfrac{\kappa^{(G)}_{x/2,\, y/2} - \min_{\kappa^{(G)}}}{\max_{\kappa^{(G)}} - \min_{\kappa^{(G)}}} \right) & \text{if } \Phi^{-1}_{x/2,\, y/2} \neq \{\varnothing\} \\ 255 \cdot \left(\dfrac{u_{xy} - \min_u}{\max_u - \min_u} \right) & \text{if } \Phi^{-1}_{x/2,\, y/2} = \{\varnothing\} \end{cases},$$

$$b_{xy} = \begin{cases} 255 \cdot \left(\dfrac{\kappa^{(B)}_{x/2,\, y/2} - \min_{\kappa^{(B)}}}{\max_{\kappa^{(B)}} - \min_{\kappa^{(B)}}} \right) & \text{if } \Phi^{-1}_{x/2,\, y/2} \neq \{\varnothing\} \\ 255 \cdot \left(\dfrac{u_{xy} - \min_u}{\max_u - \min_u} \right) & \text{if } \Phi^{-1}_{x/2,\, y/2} = \{\varnothing\} \end{cases}$$

where $\kappa^{(R)}_{x,\, y}$ denotes the number of links of the resources represented on a given pixel, thus

$$\kappa^{(R)}_{x,\, y} = \sum_{a_i \in \Phi^{-1}_{x,\, y}} \gamma^{G}_{a_i},$$

$\kappa^{(G)}_{x,\, y}$ the number of resources represented on a single pixel:

$$\kappa^{(G)}_{x,\, y} = \left| \Phi^{-1}_{x,\, y} \right|,$$

and $\kappa^{(B)}_{x,\, y}$ an empirically determined value corresponding to the directory level found in the URLs.

4. Results

This chapter presents the implementation of the prototype and the experiments that have been done with it. The goal of the experiments was not the construction of a perfect tool, rather the aim was that the prototype would demonstrate the technical feasibility.

4.1 Datasets

Multiple collections of resources with their links have been gathered from the World-Wide Web. Using a modified version of the Explore [Nierstrasz, 1995] Perl script, the following datasets have been constructed:

- `heiwww`: the resources available on the Graduate Institute of International Studies World-Wide Web server.[1]
- `liawww`: the resources available on the Artificial Intelligence Laboratory World-Wide Web server.[2]
- `iamwww`: the resources available on the Institute of Computer Science and Applied Mathematics World-Wide Web server.[3]
- `isoe`: the resources available on the School of Engineering of Oensingen World-Wide Web server.[4]
- `isbiel`: the resources available on the School of Engineering of Biel/Bienne World-Wide Web server.[5]
- `tecfa`: the resources available on the *Technologies de Formation et Apprentissage* World-Wide Web server.[6]
- `tecfamoo`: the resources available on the *Technologies de Formation et Apprentissage* MOO.[7]
- `depth`: the resources discovered through a limited depth-first search.[8]
- `breadth`: the resources discovered through a limited breadth-first search.[9]
- `ch`: the resources discovered through a limited breadth-first search[10] with priority given to unvisited World-Wide Web servers.
- `ops`: the resources discovered through a breadth-first search[11] with a number of requests per World-Wide Web server restricted to one.
- `unine`: the resources available on the various World-Wide Web servers of the University of Neuchâtel.[12]

1. Starting at <URL:http://heiwww.unige.ch/>.
2. Starting at <URL:http://liawww.epfl.ch/>.
3. Starting at <URL:http://iamwww.unibe.ch/>.
4. Starting at <URL:http://www.isoe.ch/>.
5. Starting at <URL:http://www.isbiel.ch/>.
6. Starting at <URL:http://tecfa.unige.ch/>.
7. Starting at <URL:http://tecfamoo.unige.ch/>.
8. Starting at <URL:http://heiwww.unige.ch/swizterland/>.
9. Starting at <URL:http://heiwww.unige.ch/swizterland/>.
10. Starting at <URL:http://heiwww.unige.ch/swizterland/>.
11. Starting at <URL:http://heiwww.unige.ch/swizterland/>.
12. Starting at <URL:http://www.unine.ch/>.

The adjacency and distance matrices have then been built using an implementation of the Floyd's algorithm [Sedgewick, 1989].

Some statistical values for these datasets are presented in annex B. "Statistics of the datasets" on page 59.

4.2 Maps

From the above datasets, the dimensionality reduction and the representation of the landscape have been made. The self-organizing map program package [Kohonen et al., 1995] has been improved in various areas and used to train the neural network and construct the unified matrix. Figure 13 shows the maps resulting from completion of these tasks. In the maps, small crosses represent locations and grayscale the landscape.

The self-organizing maps used are composed of an input layer of a number of units equal to the number of resources. The output is made over a grid lattice with 64 by 64 units.

The training of the self-organizing maps was the most computational intensive task. For the large datasets, it takes more than 24 hours of computation to complete. Hopefully, a parallel implementation of the learning algorithm is possible (see section 5.7 "Parallel implementation" on page 34).

Having completed the self-organizing maps training process, each resource was presented to the network, which responded with a winning neuron, the location of the resources on the map.

To allow interaction with these maps, they have been made available on-line on the World-Wide Web. Using HTML forms [Berner-Lee et al., 1995b] generated by programs that comply with the CGI (Common Gateway Interface) [Robinson, 1995] specifications [Grobe, 1995], a graphical user interface has been created. Therefore, the graphical user interface is made using dynamic hypermedia documents. Figure 14 depicts its visual aspect.

4.3 Usability

A short usability study, based upon the feedback from the first users, has been made. After empirical interpretation of some maps, it resulted that they provided an original and meaningful way to globally visualize the structure of some parts of the World-Wide Web. Although, at a local level, the ordering of neighboring resources on the maps sometimes has no obvious meaning.

The maps already revealed some information about the organization of various World-Wide Web servers. They permitted to localize the main virtually-visible actors and to interpret their interrelations.

The user interface showed that it provides an efficient one-step teleporting possibility to go back to already visited locations. Although, some works should be done to transform it into a good navigation tool.

The first users were enthusiasts at making experiments and were hopeful for the future of such a representation of cyberspace.

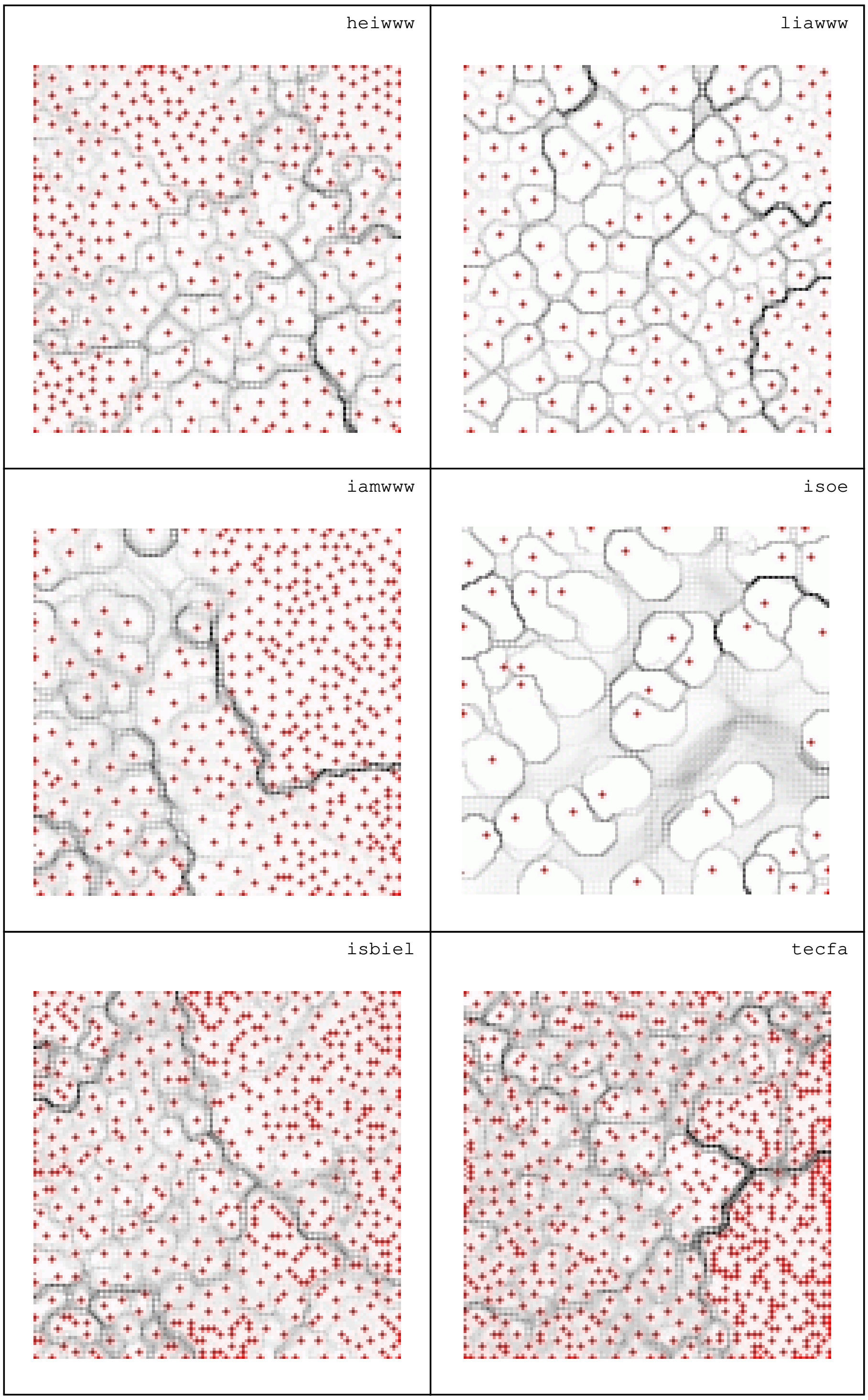


FIGURE 13. **The various maps.**

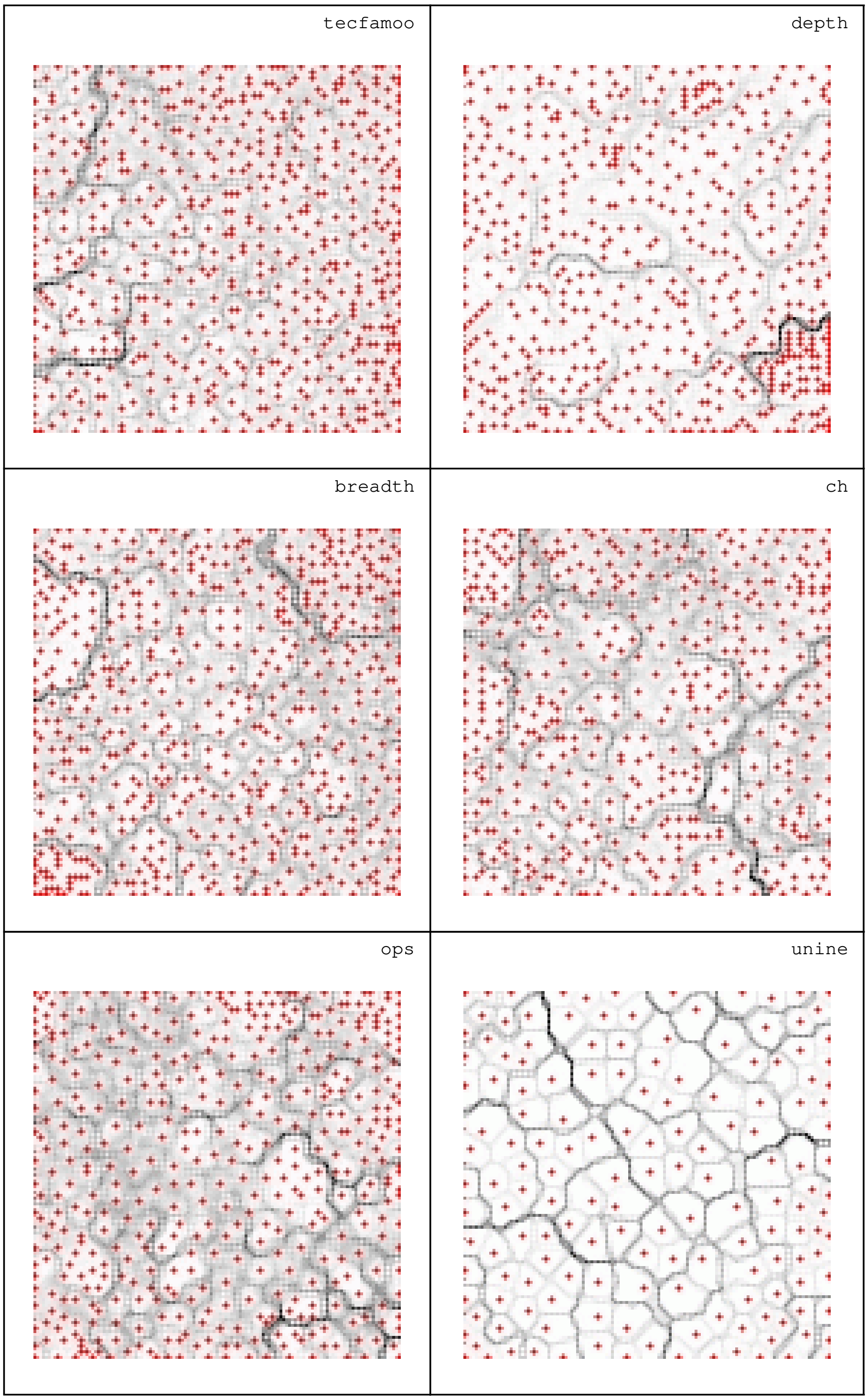


FIGURE 13. **The various maps.**

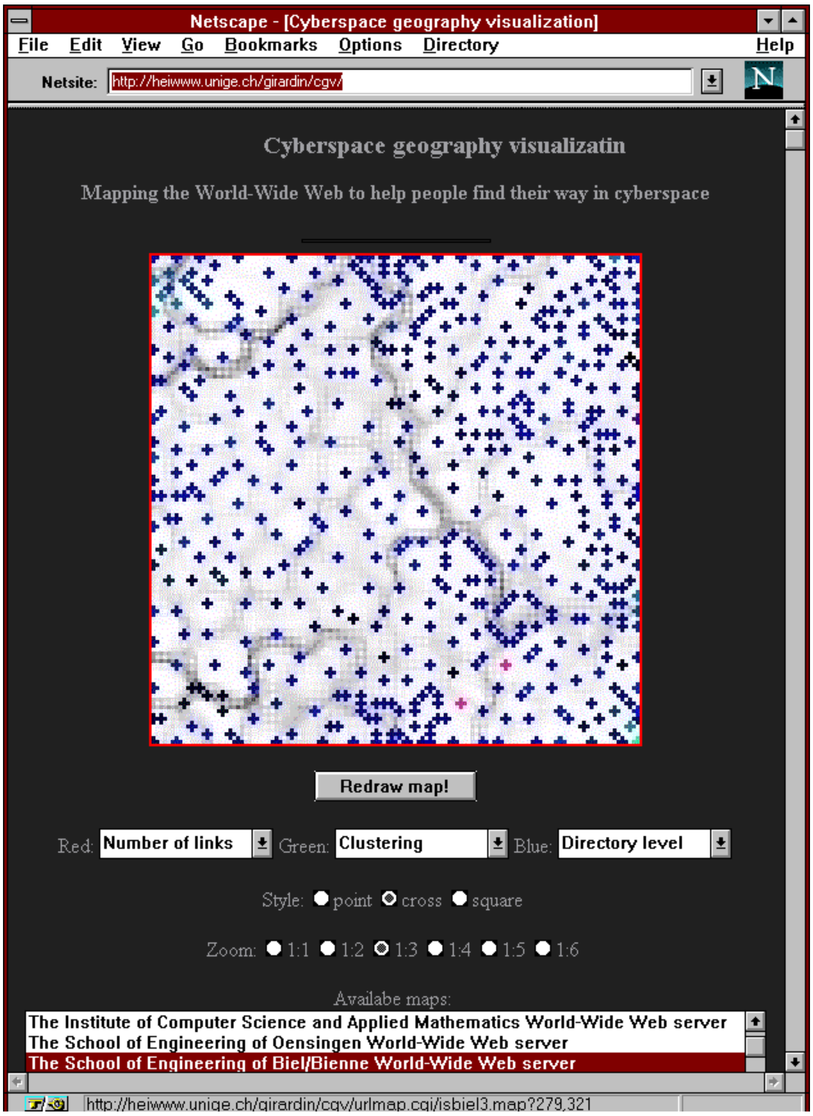


FIGURE 14. **The graphical user interface.**

5. Possible enhancements

Some ideas for future plans are mentioned below. The list is obviously not exhaustive.

5.1 Improved search strategy

To improve the search in the graph representing cyberspace, more clever exploration strategies, like a priority-first search or the Kruskal method [Sedgewick, 1989], could be done. This would require to have the search made on a weighted graph, i.e. a ranking scheme for the links contained in the World-Wide Web must be developed. To make the search faster, the use of collaborative software agents [Lashkari et al., 1994] can radically improve the time required for the information gathering. Partial information gathering, founded on empirical knowledge, could also be envisaged.

5.2 Use of different metrics

In this work, a simple arbitrary metric is defined to immerse the resources in a high dimensional space. Based on other dissimilarity coefficients [Gower and Legendre, 1986] or empirical knowledge, a smarter way to calculate the distances between elements could eventually result in a more accurate interpretation of the relationship among resources.

5.3 Quality of the mapping

The quality of the mapping has been tested empirically only. Various methods can be used to evaluate numerically how well the mapping adhere to the monotonicity constraint. These methods includes the Kruskal stress function [Kruskal, 1964], the Spearman rank correlation coefficient [Li et al., 1995] and the procrustes analysis [Cox et al., 1994]. To analyse the quality graphically, scatter-plot diagrams can give a visual display of the data correlation.

5.4 Improved visualization

The representation of the resources and of the landscape on the visualization media can be ameliorated in many ways. Making iconic or textual representation of some important locations is certainly a way to improve human-understanding. The use of other values to characterize a resources can also make some important locations emerge. As an alternative to the representation of the landscape, displaying the directions of the weight vectors can result in something similar to the representation of the flow of water in a ocean. Having the locations mapped on a globe can provide a more interesting display, giving the possibility to see a better correspondence with the geography of our planet.

5.5 Three-dimensional visualization

A dimensionality reduction to a space with three dimensions can be achieved using the self-organizing maps algorithm. The problem resides in the way to visualize it. A good direction to follow is [Wood et al., 1995].

5.6 Improved user interface

The user interface of the actual prototype provides limited interaction. To give the possibility to navigate with, at the same time, having the current location shown on the map is necessary to transform the tool into a geographic positioning system, leading to improved navigability. The drawing of routes between given locations could also provide some interesting results.

5.7 Parallel implementation

As mentioned in this paper, the most computational intensive task is the dimensionality reduction made by the self-organizing maps. Fortunately, a decomposition into small independent tasks is possible and a parallel implementation can easily be developed. The design of a possible architecture for SIMD (Single Instruction stream, Multiple Data stream) computers is explained in [Ultsch et al., 1992]. Some other parallel-computer implementations are referenced in [Kohonen, 1995].

6. Conclusion

Throughout this work, the visualization of cyberspace “common mental geography” has been investigated and illustrated with the World-Wide Web. As classical geography is based upon the interaction of atoms, the geography of cyberspace has been built upon the relationships between resources.

Using the topology of the World-Wide Web, a metric as been defined. For this purpose, modelling through graphs has been made and the distances between resources were defined as the shortest path in the graph. This enabled the possibility of representing each resource as a point in a high-dimensional space.

To permit a display of the interactions of the resources placed in a space of high dimensionality, a transformation has been required. This transformation has been made by defining a mapping of the resources onto low-dimensional visualization media, typically lattices with two dimensions. The constraint made on the mapping was to keep the monotonicity. Thus, only the rank order of the distances are preserved, protecting the most important features of the system.

To perform this non-linear dimensionality reduction, the self-organizing maps algorithm has been shown to provide the best results. The self-organizing maps method is an unsupervised neural network model that produce topology preserving maps.

Although, a model for topologically-organized maps was not sufficient because the order of similarity between resources cannot be visualized. To address with this problem, a model for representing the reliefs of self-organizing maps, the unified matrix method, was followed. By analysing the weight vectors of the self-organizing maps, a representation of the landscape became possible. It was then possible to interpret the components of this landscape as being the equivalent to mountains, ravines and valleys.

Based upon the above mentioned methods, an experimental prototype has been built. This included a software agent to gather the information, a program to immerse resources in a high dimensional space, the computation of the self-organizing maps to produce maps of low dimensionality, the creation of the unified matrix to represent the reliefs, and the development of a graphical user interface to permit the visualization of the resulting geographical maps and to give the possibility to access directly the resources behind the maps.

Since the prototype was only of academic purpose, various improvements were sketched to improve its usability. The methods used were made scalable in their spirit and therefore taking scalability into account was also possible.

The results, made available in the World-Wide Web, were shown to provide an original way to improve cyberspace navigability and to address the lost-in-cyberspace syndrome problem. It is thus encouraged that further research be done in this direction.

Glossary

CGI	Common Gateway Interface. The standard interface that World-Wide Web clients and servers use to communicate data for the creation of interactive applications.
Cyberspace	The concept of navigation through a space of electronic data, and of control which is achieved by manipulating those data.
Dimension	A measurable spatial extent.
Distance	The extent of space between two objects.
HTML	HyperText Markup Language. The standard language used for creating hypermedia documents within the World-Wide Web.
HTTP	HyperText Transfer Protocol. The standard protocol that World-Wide Web clients and servers use to communicate.
Internet	The Internet is a world-wide network of networks.
Geodesic	The shortest line between two points on any mathematically defined surface.
Geography	The topographical features of any complex entity.
Graph	A representation that exhibits a relationship between two sets as a set of points having coordinates determined by the relationship.
Hypermedia	The same as hypertext with the difference that it can contain links from and to multimedia documents.
Hypertext	Text documents containing connections within the text to other documents.
Lost-in-cyberspace	In a state where further cyberspace navigability cannot be pursued because too few or too many directions can be followed.
Map	The correspondence of one or more elements in one set to one or more elements in the same set or another set.
Mapping	A rule of correspondence established between sets that associates each element of a set with an element in the same or another set.
Metric	A function defined for a coordinate system such that the distance between any two points in that system may be determined from their coordinates.

MIME	Multipurpose Internet Mail Extensions.
MOO	Multi-user dungeons/dimensions, Object Oriented. A system that can be characterized as a multi-user, interactive and programmable virtual environment.
Morphism	The condition or quality of having a specified form.
Neural network	A system that exhibits the kind of biological computation performed in the brain.
NNTP	Network News Transport Protocol.
Pixel	The smallest image-forming unit of a picture. Contraction of picture and element.
RGB color model	A model that decomposes color into channels of red, green and blue intensity.
Self-organizing maps	A particular kind of neural network that performs a topology preserving mapping.
SGML	Standard Generalized Markup Language. A standard language to specify the structure of documents.
Somatotopic map	An associative area of the brain that performs a topology preserving mapping of sense organs on the somatosensory cortex.
Space	A set of elements or points satisfying specified geometric postulates.
Structure	The interrelation or arrangement of parts in a complex entity.
TCP	Transmission Control Protocol. A connection-oriented protocol that provides a reliable by stream for a user process.
Tonotopic map	An associative area of the brain that performs a topology preserving mapping of acoustic frequencies on the auditory cortex.
Topology	The properties of geometric figures.
Unified matrix method	A method to represent the relief of self-organizing maps.
URL	Uniform Resource Locator. A standardized way of identifying different documents, media, and network services on the World-Wide Web.

Visualization	The process of transforming information into a visual form, enabling users to observe the information.
VRML	Virtual Reality Markup Language. A standard language for describing three-dimensional hypermedia objects.
World-Wide Web	A hypermedia system running on top of the Internet.
World-Wide Web project	An initiative to create a universal, hypermedia-based method of access to information.

Bibliography

Philosophy

Philosophical discussions about issues of the objective contents of thought and of patterns of pure information can be found in [Popper, 1979][White, 1988][Penrose, 1989].

Cyberspace

The term cyberspace was coined in [Gibson, 1984]. The same author continued to describe his vision in [Gibson, 1986][Gibson, 1988][Gibson and Sterling, 1991][Gibson, 1994].The reference is certainly [Benedikt, 1991]. Interesting views on this subject are contained in [Saco, 1994][Åm, 1994][Pesce, 1994]. [Hamit, 1993] gives an introduction with several examples. An excellent introduction to MOO technology, the closest implementation of the concept of cyberspace, can be found in [Schneider et al., 1995].

World-Wide Web

An introduction to the World-Wide Web project can be found in [Hughes, 1994]. Frequently asked questions are answered in [Boutell, 1995a]. A description of the actual standardization is in [Berner-Lee et al., 1995a][Berner-Lee et al., 1995b]. The latest major developments were collected in [Kroemker, 1995][Holzapfel, 1995][Cailliau et al., 1994]. A cognitive model for structuring the World-Wide Web can be found in [Eklund, 1995].

Software agents

A discussion on issues of software agents for the World-Wide Web are discussed in [Koster, 1995b][Eichmann, 1994]. A concise comparison is provided in [Selberg and Etzioni, 1995]. A collaborative software agents model is presented in [Lashkari et al., 1994]. Suggestions for the construction of ethical World-Wide Web agents are in [Koster, 1995a][Koster, 1995c].

Measurement

The foundations of measurement are contained in [Krantz et al., 1971][Suppes et al., 1989][Luce et al., 1990]. Other related documents are [Gower and Legendre, 1986][Berka, 1983][Humphreys, 1994][McCarty, 1988].

Geography

Various issues dealing with geography are reminded in [Dollfus, 1970][Clozier, 1949][Ritter, 1971][George, 1962][Célérier, 1961].

System thinking

To model complex systems, system thinking is an important help. This approach is explained in [Gander, 1993][Morin, 1994][Le Moigne, 1990][Rosnay, 1975][Morin, 1977].

Mathematical modelling

Modelmaking of systems through formal mathematical representation is exposed in [Casti, 1992a][Casti, 1992b].

Social network analysis

An introduction to the analysis of social networks can be found in [Scott, 1991]. An extensive presentation is given in [Wassermand and Faust, 1994].

Graph theory

Introduction to the graph theory can be found in [Bollobás, 1990][Gondran and Minoux, 1995]. Of related interest are [Vosselman, 1992][Grötschel et al., 1993].

Graph drawing

A complete annotated bibliography about algorithms for drawing graphs is presented in [Batista et al., 1994]. A review of current advances can be found in [Garg and Tamassia, 1994]. Some developments dealing with graph drawing are explained in [Henry and Hudson, 1990][Cruz and Tamassia, 1994a][Cruz and Tamassia, 1994b][Eades, 1984][Chrobak et al., 1994][Fraysseix et al., 1990][Kant, 1993b][Schnyder, 1990]

Placement

An excellent review of VLSI cell placement techniques is [Shahookar and Mazumder, 1991]. A general introduction dealing with partitioning, assignment and placement can be found in [Zobrist, 1994][Goto et al., 1986]. Combinatorial algorithms for integrated circuit layout are described in [Lengauer, 1990].

Multidimensional scaling

The analysis of multidimensional datasets is covered in [Auray et al., 1990a][Auray et al., 1990b][Auray et al., 1990c][Auray et al., 1990d]. Principal component analysis is presented in depth in [Jolliffe, 1986]. Metric and non-metric multidimensional scaling are explained in [Cox et al., 1994][Kruskal and Wish, 1981][Kruskal, 1964]. Non-metric methods are compared in [Li et al., 1995].

Neural networks

A good introduction to the theory of neural computation is located in [Müller and Wyler, 1994][Blayo, 1995][Hertz et al., 1991]. General presentations can be found in [Carling, 1992][Davalo et al., 1990][Freeman, 1994].

Self-organizing maps

The reference book about self-organizing maps is certainly [Kohonen, 1995]. Other presentations can be found using references located into section “Neural networks”. Various applications can be found in [Wyler, 1994][Ultsch, 1993][Ultsch et al, 1994]. A good starting point for proving the convergence of the self-organizing maps algorithm is certainly [Cohen et al., 1987]. Other possible directions toward this goal are given in [Kohonen, 1995]. The unified matrix method for exploratory data analysis was first introduced in [Ultsch and Siemon, 1989]. An alternative to this method is given in [Kraaijveld et al., 1993]. Comparison with statistical clustering methods is presented in [Ultsch, 1995][Ultsch and Vetter, 1994].

Clustering techniques

Various clustering methods are described in [Belaïd and Belaïd, 1992]. Current developments can be found in [Dasarathy, 1990][Backer and Gelsema, 1992][Freeman, 1990].

Optimization techniques

An introduction to mathematical optimization can be found in [Computational Science Education Project, 1995]. A reading list about combinatorial optimization is located in [Borchers, 1994]. A good method for combinatorial optimization, simulated annealing, is reviewed in [Larrhoven and Aarts, 1988][Ingber, 1995][Ingber, 1993][Ingber, 1989]. Evolutionary computation strategies can also be used for optimization; some relevant documents are [Goldberg, 1994][Soucek et al., 1992][Bäck et al., 1991][Bramlette, 1991]. A comparative study between simulated annealing and evolution strategy is presented in [Groot et al., 1990]. Various approaches to large-scale optimization are presented in [Coleman, 1991][Conn et al., 1994][Karmat, 1993]. Of related interest are [Gent and Walsh, 1993][Minton et al., 1994]. Other classical algorithms are described in [Sedgewick, 1989][Press et al., 1994]. A guide to optimization software can be found in [Moré and Wright, 1993].

Parallel computing

Parallel implementation of the self-organizing map algorithm is presented in [Ultsch et al., 1992]. Designing efficient algorithms for parallel computers is introduced in [Quinn, 1994]. A good portable language for implementing parallel algorithms is [Droux, 1995].

Computational complexity

An excellent guide to the theory of NP-completeness is [Garey et al., 1979]. Combinatorial reasoning is analysed in [Tucker, 1995]. Other documents related to computational complexity are [Bonuccelli et al., 1994][Brauer et al., 1984][Miller and Orlin, 1985].

Visualization

Advances in visualization technologies are reviewed in [Rosenblum et al., 1994][Schneiderman, 1995]. An approach to the visualization of complex systems is presented in [Hendley and Drew, 1995][Drew et al., 1995] and used to visualize the World-Wide Web in [Wood et al., 1995]. Another environment for visualizing the World-Wide Web is presented in [Kent and Neuss, 1994]. Good introductions to computer graphics are [Burger and Gillies, 1989][Rogers, 1988]. Interesting ideas to improve the visualization can be found in [Grossberg, 1983][Haber, 1983][Kant, 1993a].

References

[Åm, 1994] Åm, Onar. **Cyberspace and the structure of knowledge**. 1994. <URL:http:/ /www.hsr.no/~onar/Ess/ Cyberspace_and_the_Structure_of_Knowledge.html>

[Auray *et al.*, 1990a] Auray, J. P.; Duru, G., and Zighed, A. ***Analyse de s données multidimensionnelles: les méthodes de structuration***. Lyon: Alexandre Lacassagne; 1990. ISBN: 2-905972-21-1.

[Auray *et al.*, 1990b] Auray, J. P.; Duru, G., and Zighed, A. ***Analyse des données multidimensionnelles: les méthodes de description***. Lyon: Alexandre Lacassagne; 1990. ISBN: 2-905972-13-0.

[Auray *et al.*, 1990c] Auray, J. P.; Duru, G., and Zighed, A. ***Analyse des données multidimensionnelles: les méthodes d'explication***. Lyon: Alexandre Lacassagne; 1991. ISBN: 2-905972-24-6.

[Auray *et al.*, 1990d] Auray, J. P.; Duru, G., and Zighed, A. ***Analyse des données multidimensionnelles: aspects méthodologiques***. Lyon: Alexandre Lacassagne; 1991. ISBN: 2-905972-26-2.

[Bäck et al., 1991] Bäck, Thomas; Hoffmeister, Frank, and Schwefel, Hans-Paul. **A survey of evolution strategies**. Belew, Richard K. and Lashon B. Booker, editors. Proceedings of the fourth International Conference on Genetic Algortihms; 1991 July 13; University of California, San Diego. San Mateo: Morgan Kaufmann; 1991; ISBN: 1-55860-208-9.

[Backer and Gelsema, 1992] Backer, E. and Gelsema, E.S., general chairmen. **Proceedings of the 11th IAPR International Conference on Pattern Recognition**; The Hague; 1992 August 30-September 3.

[Batista *et al.*, 1994] Battista, Giuseppe Di; Tamassia, Roberto; Eades, Peter; Tollis, Ioannis G. **Algorithms for drawing graphs: an annotated bibliography**. 1994 June. <URL:ftp://wilma.cs.brown.edu/pub/papers/compgeo/gdbiblio.ps.Z>.

[Benedikt, 1991] Benedikt, Michael. **Cyberspace: first steps**. Cambridge: The MIT Press; 1991. ISBN: 0-262-02327-X.

[Belaïd and Belaïd, 1992] Belaïd, Abdel and Belaïd, Yolande. ***Reconnaissance des formes: méthodes et applications***. Paris: InterEditions; 1992

[Berka, 1983] Berka, Karel. **Measurement: its concepts, theories and problems**. Dordrecht; Boston; London: Reidel; 1983

[Berner-Lee *et al.*,

1995a] Berner-Lee, T.; Fieding, R. T., and Frystyk Nielsen, H. **Hypertext Transfer Protocol - HTTP/1.0**; 1995. <URL:http://www.w3.org/pub/WWW/Protocols/HTTP1.0/draft-ietf-http-spec.html>.

[Berner-Lee *et al.*, 1995b] Berner-Lee, T.; Connolly D. **Hypertext Markup Language - 2.0**; 1995 September 22. <URL:http://www.w3.org/pub/WWW/MarkUp/html-spec/html-spec_toc.html>.

[Blayo, 1995] Blayo, François. ***Réseaux neuronaux***; 1995

[Bollobás, 1990] Bollobás, Béla. **Graph theory: an introductory course**. 3 ed. New York: Springer; 1990. Claude Fuhrer; ISBN: 0-387-90399-2.

[Bonuccelli *et al.*, 1994] Bonuccelli, M.; Crescenzi, P.; Petreschi, R.; editors. **Algorithms and complexity**. Berlin: Springer; 1994. ISBN: 3-540-5711-0.

[Borchers, 1994] Borchers, Brian. **A reading list in combinatorial optimization**. 1994 April 21.

[Boutell, 1995a] Boutell, Thomas. **World Wide Web Frequently Asked Questions**. 1995. <URL:http://sunsite.unc.edu/boutell/faq/www_faq.html>

[Boutell, 1995b] Boutell, Thomas. **cgic: an ANSI C library for CGI programming**; 1995. <URL:http://sunsite.unc.edu/boutell/cgic/cgic.html>.

[Bramlette, 1991] Bramlette, Mark F. **Initialization, mutation and selection methods in genetic algorithms for function optimization**. Belew, Richard K. and Booker, Lashon B., editor. Proceedings of the fourth International Conference on Genetic Algorithms; 1991 Jul 13; University of California, San Diego. San Mateo: Morgan Kaufmann; 1991; 100-107. ISBN: 1-55860-208-9.

[Brauer *et al.*, 1984] Brauer, W.; Rozenberg, G.; Salomaa, A., editors. **Graph algorithms and NP-completeness**. Berlin: Springer; 1984. ISBN: 3-540-13541-X.

[Burger and Gillies, 1989] Burger, Peter and Gillies, Duncan. **Interactive computer graphics**. Wokingham: Addison-Wesley; 1989. ISBN: 0-201-17439-1.

[Cailliau *et al.*, 1994] Cailliau, R.; Nierstrasz, O., and Ruggier, M., editors. **Proceedings of the first International World-Wide Web Conference**; 1994 May 25; Geneva.

[Carling, 1992] Carling, Alison. **Introducing neural networks**. Wilmslow: Sigma; 1992. ISBN: 1-85058-174-6.

[Casti, 1992a] Casti, John L. **Reality rules: picturing the world in mathematics: the fundamentals**. New York; Toronto: Wiley; 1992. ISBN: 0-471-57021-4.

[Casti, 1992b] Casti, John L. **Reality rules: picturing the world in mathematics: the frontier**. New York; Toronto: Wiley; 1992. ISBN: 0-471-57798-7.

[Célérier, 1961] Célérier, Pierre. ***Géopolitique et géostratégie***. Paris: Presses Universitaires de France; 1961.

[Chrobak *et al.*, 1994] Chrobak, Marek and Nakano, Shin-ichi. **Minimum-width grid drawings of plane graphs**; 1994.

[Clozier, 1949] Clozier, René. ***Les étapes de la géographie***. Paris: Presses Universitaires de France; 1949.

[Cohen *et al.*, 1987] Cohen, Michael A. and Grossberg, Stephen. **Absolute stability of global pattern formation and parallel memory storage by competitive neural networks** Grossberg, Stephen, editor. The adaptive brain. Amsterdam: Elsevier; 1987: 288-308. ISBN: 0-444-70117-6.

[Coleman, 1991] Coleman, Thomas F. **Large-scale numerical optimization: introduction and overview**; 1991 September. <URL:http://cs-tr.cs.corness.edu/TR/CORNELLCS:TR91-1236/Print>.

[Computational Science Education Project, 1995] Computational Science Education Project. **Mathematical optimization**. 1995. <URL:http://csep1.phy.ornl.gov/mo/mo.html>.

[Conn *et al.*, 1994] Conn, A. R.; Gould, N. I. M., and Toint, Ph. L. **Large-scale nonlinear constrained optimization: a current survey**. 1994 February 1. <URL:ftp://seamus.cc.rl.ac.uk/pub/reports/cgtCERFACS9403.ps.Z>.

[Connoly, 1995a] Connoly, Daniel W. **HyperText Markup Language (HTML): working and background materials**. 1995. <URL:http://www.w3.org/pub/WWW/MarkUp/MarkUp.html>

[Connoly, 1995b] Connoly, Daniel W. **UR* and the names and addresses of WWW objects**. 1995. <URL:http://www.w3.org/pub/WWW/Addressing/Addressing.html>.

[Cox *et al.*, 1994] Cox, Trevor F. and Cox, Michael A. A. **Multidimensional scaling**. London: Chapman & Hall; 1994. (Monographs on statistics and applied probability; 59). ISBN: 0-412-49120-6.

[Cruz and Tamassia, 1994a] Cruz, Isabel F.; Tamassia, Roberto. **How to visualize a graph: specification and algorithms: algorithmic approach**. 1994. <URL:ftp://ftp.cs.brown.edu/pub/papers/compgeo/draw-tutorial-1.ps.Z>.

[Cruz and Tamassia, 1994b] Cruz, Isabel F.; Tamassia, Roberto. **How to visualize a graph: specification and algorithms: declarative approach**. 1994. <URL:ftp://ftp.cs.brown.edu/pub/papers/compgeo/draw-tutorial-2.ps.Z>.

[Dasarathy, 1990] Dasarathy, Belur V. **Nearest neighbor(NN) norms: NN pattern classification techniques**. Washington; Brussels; Tokyo: IEEE Computer Society Press. 1990.

[Davalo *et al.*, 1990] Davalo, Eric and Naïm, Patrick. ***Des réseaux de neurones***. 2 ed. Paris: Eyrolles; 1990. ISBN: 2-212-08137-5.

[Dollfus, 1970] Dollfus, Olivier. ***L'espace géographique***. Paris: Presses Universitaire de France; 1970.

[Drew *et al.*, 1995] Drew, Nick and Hendley, Bob. **Visualising complex interacting systems**; 1995. <URL:http://www.cs.bham.ac.uk/~nsd/CHI95/nsd_bdy.html>.

[Droux, 1995] Droux, Nicolas. **PPC: Portable Parallel C**. 1995. <URL:http://www.isbiel.ch/I/Projects/ppc/>.

[Eades, 1984] Eades, Peter. **A heuristic for graph drawing** Stanton. Ralph G., editor. Congressus numerantium. Winnipeg: Utilitas Mathematica; 1984 May: 149-160. ISBN: 0-919628-42-7.

[Eichmann, 1994] Eichmann, David. **Ethical web agents**; 1994. <URL:http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Agents/eichmann.ethical/eichmann.html>.

[Eklund, 1995] Eklund, John. **Cognitive models for structuring hypermedia and implications for learning from the world-wide web**; 1995. <URL:http://www.scu.edu.au/ausweb95/papers/hypertext/eklund/index.html>.

[Fraysseix *et al.*, 1990] Fraysseix, H. de; Pach, J., and Pollack, R. **How to draw a planar graph on a grid**. Combinatorica: Springer; 1990: 41-51.

[Freeman, 1990] Freeman, Herbert. **Proceedings of the 10th International Conference on Pattern Recognition**; Atlantic City; 1990 June 16-21.

[Freeman, 1994] Freeman, James A. **Simulating neural networks with Mathematica**. Reading: Addison-Wesley; 1994. ISBN: 0-201-56629-X.

[Gander, 1993] Gander, Jean-Gabriel. ***Systems engineering: integriertes Denken, Konzipieren und Realisieren von lebensfähigen System***. Bern: Ingenieurshule Bern HTL; 1993.

[Garey *et al.*, 1979] Garey, Michael R. and Johnson, David S. **Computers and intractability: a guide to the theory of NP-completeness**. New York; Oxford: Freeman; 1979. ISBN: 0-7167-1044-7.

[Garg and Tamassia, 1994] Garg, Ashim and Tamassia, Roberto. **Advances in graph drawing**. Bonuccelli, M.; Crescenzi, P., and Petreschi, R., editor. Algorithms and complexity; 1994 Feb; Rome. Berlin; Heidelberg: Springer; 1994: 13-21. ISBN: 3-540-57811-0.

[Gent and Walsh, 1993] Gent, Ian P. and Walsh, Toby. **An empirical analysis of search in GSAT**. 1993. <URL:http://ic-www.arc.nasa.gov/ic/jair-www/volume1/gent93a.ps>.

[George, 1962] George, Pierre. ***Géographie industrielle du monde***. Paris: Presses Universitaires de France; 1962.

[Gibson, 1984] Gibson, William. **Neuromancer**. New York: Ace; 1984. ISBN: 0-441-56959-5.

[Gibson, 1986] Gibson, William. **Count zero**. London: HarperCollins; 1986. ISBN: 0-586-07121-0.

[Gibson, 1988] Gibson, William. **Mona Lisa overdrive**. New York: Bantam. 1988.

[Gibson, 1994] Gibson, William. **Virtual light**. London: Penguin; 1994. ISBN: 0-14-015772-7.

[Gibson and Sterling, 1991] Gibson, William and Sterling, Bruce. **The difference engine**. New York: Bantam; 1991

[Goldberg, 1994] Goldberg, David E. **Algorithmes génétiques: exploration, optimisation et apprentissage automatique**. Paris: Addison-Wesley; 1994. ISBN: 2-87908-0054-1.

[Gondran and Minoux, 1995] Gondran, Michel and Minoux, Michel. **Graphes et algorithmes**. 3 ed. Paris: Eyrolles; 1995.

[Goto *et al.*, 1986] Goto, Satoshi and Matsuda, Tsuneo. **Partitioning, assignment and placement**. Ohtsuki, T., editor. Layout design and verification. North-Holland: Elsevier; 1986

[Gower and Legendre, 1986] Gower, J. C. and Legendre, P. **Metric and Euclidean properties of dissimilarity coefficients**. Journal of classification. New York: Springer; 1986: 5-48.

[Grobe, 1995] Grobe, Michael. **Instantaneous introduction to CGI scripts and HTML forms**. 1995. <URL:http.//kuhttp.cc.ukans.edu/info/forms/forms-intro.html>.

[Groot *et al.*, 1990] Groot, Class de; Würtz, Diethelm, and Hoffmann, Karl Heinz. **Algorithms: simulated annealing and evolution strategy - a comparative study**. Zurich: 1990 September.

[Grossberg, 1983] Grossberg, Stephen. **The quantized geometry of visual space: the coherent computation of depth, form, and lightness**. Harnad, Stevan, editor. The behavioral and brain sciences. Cambridge: Cambridge University; 1983: 625-692.

[Grötschel et al., 1993] Grötschel, Martin; Lovász, László, and Schrijver, Alexander. **Geometric algorithms and combinatorial optimization**. Second corrected edition. Berlin: Springer; 1993. ISBN: 3-540-56740-2.

[Haber, 1983] Haber, Ralph Norman. **The impending demise of the icon: a critique of the concept of iconic storage in visual information processing**. Harnad, Stevan, editor. The bahavioral and brain sciences. Cambridge: Cambridge University; 1983: 1-54.

[Hamit, 1993] Hamit, Francis. **Virtual reality and exploration of cyberspace**: Sams; 1993

[Hendley and Drew, 1995] Hendley, Bob and Drew, Nick. **Visualisation of complex systems**; 1995. <URL:http://www.cs.bham.ac.uk/~nsd/HCI95/hci95.html>.

[Henry and Hudson, 1990] Henry, Tyson R. and Hudson, Scott E. **Viewing large graphs**. Tucson: The University of Arizona; 1990.

[Hertz *et al.*, 1991] Hertz, John; Krogh, Anders; Palmer, Richard D. **Introduction to the theory of neural computation**. Reading: Addison-Wesley; 1991. ISBN: 0-201-50395-6.

[Holzapfel, 1995] Holzapfel, Roland, editor. **Poster proceedings of the Third International World-Wide Web Conference**; 1995 April 10; Darmstadt.

[Hughes, 1994] Hughes, Kevin. **Entering the World-Wide Web: a guide to cyberspace**; 1994 May. <URL:http://www.eit.com/web/www.guide>.

[Humphreys, 1994] Humphreys, Paul. **Patrick Suppes: scientific philosopher: philosophy of physics, theory structure, and measurement theory** . Dordrecht; Boston; London: Kluwer Academic; 1994. ISBN: 0-7923-2553-2.

[Ingber, 1995] Ingber, Lester. **Adaptive simulated annealing (ASA): lessons learned**. 1995. <URL:ftp://ftp.alumni.caltech.edu/pub/ingber/asa95_lessons.ps.Z>.

[Ingber, 1993] Ingber, Lester. **Simulated annealing: practive versus theory**. 1993. <URL:ftp://ftp.alumni.caltech.edu/pub/ingber/asa93_sapvt.ps.Z>.

[Ingber, 1989] Ingber, Lester. **Very fast simulated re-annealing**. 1989. <URL:ftp://ftp.alumni.caltech.edu/pub/ingber/asa89_vfsr.ps.Z>.

[Jolliffe, 1986] Jolliffe, I. T. **Pincipal component analysis**. New York: Springer; 1986. University of Geneva; ISBN: 0-387-96269-7.

[Karmat, 1993] Kamat, Manohar P. **Structural optimization: status and promise**. Washington: American Institute of Aeronautics and Astronautics; 1993.

[Kant, 1993a] Kant, Goosen. **A more compact visibility representation**. Graph-theoretic concepts in computer science: 19th International Workshop; 1993 Jun 16; Utrecht. 1993: 411-424.

[Kant, 1993b] Kant, Goossen. **Algorithms for drawin planar graphs**; 1993

[Kent and Neuss, 1994] Kent, Robert E. and Neuss, Christian. **Creating a web analysis and visualization environment**; 1994. <URL:http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Autools/kent/kent.html>.

[Kohonen, 1995] Kohonen, Teuvo. **Self-organizing maps**. Berlin; Heidelberg; New-York: Springer; 1995. ISBN: 3-540-58600-8.

[Kohonen *et al.*, 1995] Kohonen, Teuvo; Hynninen, Jussi; Kangas, Jari, and Laaksonen, Jorma. **The self-organizing map program package**. Version 3.1. 1995 April 7. <URL:ftp://cochlea.hut.fi/pub/som_pak/som_pak-3.1.tar.Z>.

[Koster, 1995a] Koster, Martijn. **Guidelines for robot writers**; 1995. <URL:http://info.webcrawler.com/mak/projects/robots/guidelines.html>.

[Koster, 1995b] Koster, Martijn. **Robots in the web: threat or treat?**; 1995. <URL:http://info.webcrawler.com/mak/projects/robots/threat-or-treat.html>.

[Koster, 1995c] Koster, Martijn. **A standard for robot exclusion**; 1995. <URL:http://info.webcrawler.com/mak/projects/robots/norobots.html>.

[Kraaijveld et al., 1993] Kraaijveld, Martin A.; Mao, Jiangchang, and Jain, Anil K. **A non-linear projection method based on Kohonen's topology preserving maps**. IEEE Computer Society, editor. Proceedings of the 11th IAPR International Conference on Pattern Recognition; 1993 August 3; The Hague. Washington; Brussels; Tokyo: IEEE; 1992 ISBN: 0-8186-2915-0.

[Krantz *et al.*, 1971] Krantz, David H.; Luce, R. Duncan; Suppes, Patrick, and Tversky, Amos. **Foundations of measurement: additive and polynomial representations**. San Diego; London: Academic Press; 1971. University of Fribourg; ISBN: 0-12-425401-2.

[Kroemker, 1995] Kroemker, D., editor. **Proceedings of the Third International World-Wide Web Conference**; 1995 April 10; Darmstadt. Amsterdam: Elsevier; 1995. <URL:http://www.elsevier.nl/>.

[Kruskal, 1964] Kruska, Joseph B. **Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis**. Adkins, Dorothy C. and Horst, Paul, editors. Psychometrika; 1964 March.

[Kruskal and Wish, 1981] Kruskal, Joseph B. and Wish, Myron. **Multidimensional scaling**. Beverly Hills; London: Sage Publications; 1981. (Quantitative applications in the social sciences; 07-011). ISBN: 0-8039-0940-3.

[Larrhoven and Aarts, 1988] Larrhoven, Peter J. M. van and Aarts, Emile H. L. **Simulated annealing: theory and applications**. Reprint with corrections ed. Dordrecht: D. Reidel; 1988. ISBN: 90-277-2513-6.

[Lashkari et al., 1994] Lashkari, Yezdi; Metral, Max, and Maes, Pattie. **Collaborative interface agents**; 1994. <URL:ftp://media-lab.media.mit.edu/pub/agents/interface-agents/coll-agents.ps.Z>.

[Le Moigne, 1990] Le Moigne, Jean-Louis. ***La modélisation des systèmes complexes***. Paris: Dunod; 1990.

[Lengauer, 1990] Lengauer, Thomas. **Combinatorial algorithms for integrated circuit layout**. Stuttgart; Chichester: Teubner; Wiley; 1990. ISBN: 0-471-92838-0.

[Li *et al.*, 1995] Li, Sofianto; Vel, Olivier de, and Coomans, Danny. **Comparative performance analysis of non-linear dimensionality reduction methods**. 1995. <URL:http://www.cs.jcu.edu.au/ftp/pub/techreports/94-8.ps.gz>.

[Luce *et al.*, 1990] Luce, R. Duncan; Krantz, David H.; Suppes, Patrick, and Tversky, Amos. **Foundations of measurement: representation, axiomatization, and invariance**. San Diego; London: Academic Press; 1990. ISBN: 0-12-425403-9.

[McCarty, 1988] McCarty, George. **Topology: an introduction with application to topological groups**. New York: Dover; 1988. ISBN: 0-486-65633-0.

[Miller and Orlin, 1985] Miller, Z. and Orlin, J. B. **NP-completeness for minimizing maximum edge length in grid embeddings**. Journal of algorithms; 1985: 10-16.

[Minton *et al.*, 1994] Minton, Steven; Bresina, John, and Drummond, Mark. **Total-order and partial-order planning: a comparative analysis**. Journal of Artificial Intelligence Research; 1994: 227-262. . <URL:http://www.cs.washington.edu/research/jair/volume2/minton94a-thml/paper.html>.

[Moré and Wright, 1993] Moré, Jorge J. and Wright, Stephen J. **Optimization software guide**. Philadelphia: Society for Industrial and Applied Mathematics; 1993. (Conghran, W. M. and Grosse, Eric. Frontiers in applied mathematics; 14). ISBN: 0-89871-322-6.

[Morin, 1977] Morin, Edgar. ***La méthode***. Paris: Seuil; 1977-1991.

[Morin, 1994] Morin, Edgar. ***Introduction à la pensée complexe***. Paris: ESF.

[Müller and Wyler, 1994] Müller, Lorenz and Wyler, Kuno. ***Neuronale Netze und genetische Algorithmen***; 1994

[Nierstrasz, 1995] Nierstrasz, Oscar. **Explore: explore the WWW starting at a given URL**. 1995. <URL:http://iamwww.unibe.ch/~scg/Src/Scripts/explore>.

[Penrose, 1989] Penrose, Roger. **The emperor's new mind: concerning computers, minds, and the laws of physics**. Oxford: Oxford University Press; 1989. ISBN: 0-19-851973-7.

[Pesce, 1994] Pesce, Mark D. **Cyberspace**; 1994. <URL:http://hyperreal.com/~mpesce/www.html>.

[Popper, 1979] Popper, Karl. **Objective knowledge: an evolutionary approach**. Oxford: Oxford University Press; 1979.

[Press *et al.*, 1994] Press, William H.; Teukolsky, Saul A.; Flannery, Brian P.; Vetterling. **Numerical recipes: the art of scientific computing**. Cambridge: Cambridge University Press; 1988. ISBN: 0-521-30811-9.

[Quinn, 1994] Quinn, Michael J. **Parallel computing: theory and practice**. New York: McGraw-Hill; 1994.

[Ritter, 1971] Ritter, Jean. ***Géographie des transports***. Paris: Presses Universitaires de France; 1971.

[Robinson, 1995] Robinson, David. **The WWW Common Gateway Interface version 1.1**. 1995 May 31. <URL:http://www.wast.cam.ac.uk/~drtr/cgi-spec.html>.

[Rogers, 1988] Rogers, David F. ***Algorithmes pour l'infographie***. Auckland: McGraw-Hill; 1988. ISBN: 2-7042-1171-X.

[Rosenblum et al., 1994] Rosenblum, L.; Earnshaw, R. A.; Encarnação, J.; Hagen, H.; Kaufman, A.; Klimenko, S.; Nielson, G.; Post, F., and Thalmann, D. **Scientific visualization: advances and challenges**. London; San Diego: Academic Press; 1994. ISBN: 0-12-227742-2.

[Rosnay, 1975] Rosnay, Joël de. ***Le macroscope***: vers une vision globale. Paris: Seuil; 1975.

[Saco, 1994] Saco, Diana. **Cybercitizens: reflections on democracy an communication in the electronic frontier**. 1994 October 19.

[Schneider *et al.*, 1995] Schneider, Daniel K.; Drozdowski, Terrence M.; Glusman, Gustavo; Godard, Richard; Block, K., and Tennison, Jenifer. **The evolving TecfaMOO book - Part I: concepts**; 1995 August 31. <URL:http://tecfa.unige.ch/moo/book1/tm.html>.

[Schneiderman, 1995] Schneiderman, Ben. **User interfaces for information visualization**; 1995 March 30; Lausanne.

[Schnyder, 1990] Schnyder, Walter. **Embedding planar graphs on the grid**. Proceedings of the ACM-SIAM symposium on discrete algorithms; 1990 January 22; San Francisco. Society for Industrial and Applied Mathematics; 1990: 138-148. ISBN: 0-89871-251-3.

[Scott, 1991] Scott, John. **Social network analysis: a handbook**. London: Sage; 1991. ISBN: 0-8039-8480-4.

[Sedgewick, 1989] Sedgewick, Robert. **Algorithms**. Second ed. Reading; Don Mills: Addison-Wesley; 1989. ISBN: 0-201-06673-4.

[Selberg and Etzioni, 1995] Selberg, Erik and Etzioni, Oren. **Multi-engine search and comparison using the MetaCrawler**; 1995. <URL:http://metacrawler.cs.washington.edu:8080/paper-www4/metacrawler.html>.

[Shahookar and Mazumder, 1991] Shahookar, K. and Mazumder, P. **VLSI cell placement techniques**. ACM computing surveys; 1991 Jun: 143-220.

[Siemon and Ultsch, 1992] Siemon, H. P. and Ultsch, A. **Kohonen neural networks for exploratory data analysis**. Proceedings of the conference for information and classification, Dortmund; 1992 April.

[Soucek et al., 1992] Soucek, Branco and IRIS Group. **Dynamic, genetic, and chaotic programming: the sixth generation**. New York, Toronto: John Wiley; 1992. ISBN: 0-471-55717-X.

[Suppes *et al.*, 1989] Suppes, Patrick; Krantz, David H.; Luce, R. Duncan, and Tversky, Amos. **Foundations of measurement: geometrical, thereshold, and probalistic representations**. San Diego; London: Academic Press; 1989. ISBN: 0-12-425402-0.

[Tucker, 1995] Tucker, Alan. **Applied combinatorics**. Third ed. New York; Toronto: Wiley; 1995. ISBN: 0-471-11091-4.

[Ultsch, 1995] Ultsch, A. **Self organizing neural networks perform different from statistical k-means clustering**; 1995

[Ultsch, 1993] Ultsch, A. **Self-organizing neural networks for visualisation and classification**; 1993

[Ultsch and Siemon, 1989] Ultsch, A. and Siemon, H. P. **Exploratory data analysis: using Kohonen networks on Transputers**. 1989 December.

[Ultsch and Vetter, 1994] Ultsch, Alfred; Vetter, C. **Self-organizing-feature-maps versus statistical clustering methods: a benchmark**. 1994. <URL:http://www.mathematik.uni-marburg.de/~wina/papers/94.cluster.ps>

[Ultsch et al, 1994] Ultsch, Alfred; Korus, Dieter, and Guimaraes, Gabriela. ***Integration von neuronalen Netzen mit wissensbasierten Systemen***; 1994. <URL:http://www.mathematik.uni-marburg.de/~wina/papers/94.CoWAN.ps>.

[Ultsch *et al.*, 1992] Ultsch, A.; Guimaraes, G.; Korus, D., and Li, H. ***Selbstorganisierende neuronale Netze auf Transputern***; 1992. <URL:http://www.mathematik.uni-marburg.de/~wina/papers/92.TAT.ps>.

[Vosselman, 1992] Vosselman, George. **Relational matching**. Berlin; Heidelberg: Springer; 1992. ISBN: 3-540-55798-9.

[Wassermand and Faust, 1994] Wasserman, Stanley and Faust, Katherine. **Social network analysis: methods and applications**. Cambridge: Cambridge University Press; 1994. University of Geneva; ISBN: 0-521-38269-6.

[White, 1988] While, Leslie A. **The locus of mathematical reality: an anthropological footnote**. Newman, James R., The world of mathematics. Redmond: Tempus; 1988: 2325-2340. ISBN: 1-55615-148-9.

[Wilkinson *et al.*, 1992] Wilkinson, Leland; Hill, MaryAnn; Welna, Jeffrey P., and Birkenbeuel, Gregory K. **SYSTAT for Windows: Statistics**: SYSTAT; 1992

[Wood *et al.*, 1995] Wood, Andrew; Drew, Nick; Beale, Russell, and Hendley, Bob. **HyperSpace: web browsing with visualisation**; 1995. <URL:http://www.cs.bham.ac.uk/~amw/hyperspace/www95/>.

[Wyler, 1994] Wyler, Kuno. ***Selbstorganisierende Prozesszuordnung in einem Mehrprozessorsystem***. Bern: 1994.

[Zobrist, 1994] Zobrist, George Winston. **Routing, placement, and partitioning**. Norwood: Ablex; 1994. ISBN: 0-89391-784-2.

A. Information gathering and the World-Wide Web

The World-Wide Web is composed of resources, mostly documents, that are identified using an URL which is composed of four parts:

1. The protocol scheme, which has to be registered.
2. The fully qualified domain name of a network host, or its IP address.
3. The port number. If not specified, the default port number according to the protocol is used.
4. The rest of the locator specifies the path of the resource. It depends on the protocol used.

Most of the resources available in the World-Wide Web is transferred with the HTTP scheme. Other protocols are mainly used to give backward compatibility. A major exception is NNTP (Network News Transfer Protocol) which gives access to the Usenet News. Unfortunately, the information available through NNTP is not kept on a long term basis. Therefore, it is reasonable to restrict the search to the HTTP scheme.

The HTTP URL takes the form

```
http://<host>:<port>/<path>?<searchpart>
```

where if <port> is omitted, the port defaults to 80. It is obvious that URL containing a searchpart element can be discarded. An example of an HTTP URL is

```
http://www.w3.org/hypertext/WWW/Protocols/HTTP1.0/draft-ietf-http-spec.html
```

which is the location of the HTTP Internet draft.

According to a defined MIME (Multipurpose Internet Mail Extensions) type, a resource can be either a text, a hypertext, a picture, a sound, *etc*. Because we are interested only in information containing hyperlinks, we can focus our attention on hypertext documents which are currently only available in HTML. Note that VRML (Virtual Reality Markup Language), although in experimental testing, should soon give hyperlinks functionalities to three-dimensional immersive environments.

HTML is an application of SGML (Standard Generalized Markup Language). It permits the anchoring of parts of documents, either in textual or pictural forms, to other resources by giving their URLs. A URL can be specified in its absolute form or as a relative address. This is an example of simple HTML document with one hypertext link:

```
<HTML>
<HEAD>
<H1>This is the title</H1>
</HEAD>
<BODY>
<P>This is a paragraph with one
<A HREF="http://www.eit.com/web/www.guide/">hypertext link</P>.
</BODY>
</HTML>
```

To fetch a resource using HTTP, a connection to the specified host has to be establish over TCP (Transmission Control Protocol). The request for a resource can then be made by sending a GET command followed by the URL. A response header is returned with the information on the MIME type. As discussed before, we will limit ourselves to the text/html Content-Type. This header is normally followed by the data in the format of a MIME message body. Because no assurance is given of the existence of a resource, this fetching should be made particularly tolerant of any error.

After the HTML document has been successfully fetched, parsing its content can be made. Each discovered anchor can be put in a queue of URLs to fetch. It has to be put at the end of the queue to accomplish a breadth-first search and at the top for a depth-first search.

The next URL to fetch can then be popped from the top of the queue and this process can continue until a specified number of resources has been fetched or until the queue is empty.

Each successfully fetched URL, with all of its anchored links, can then be stored.

B. Statistics of the datasets

Some statistical values for the datasets presented in section 4.1 “Datasets” on page 27 are presented below:

Dataset	Number of resources collected	Number of collected links	Number of resolved links within the collected resources	Number of selected resources	Number of unique resolved links within the selected resources
heiwww	401	5334	1775	401	1116
liawww	155	1908	762	155	264
iamwww	5493	168829	63191	1000	1497
isoe	44	518	112	44	94
isbiel	983	25562	14474	983	3421
tecfa	4971	99691	71364	1000	3502
tecfamoo	1049	3710	3525	1049	2741
depth	1598	33717	11489	1000	1987
breadth	10001	258663	31610	1000	2926
ch	2664	1042825	10280	1000	2264
ops	10091	287545	22078	1000	1619
unine	161	1687	436	161	290

Dataset	Mean number of collected links per collected resources	Mean number of unique resolved links per selected resources	Maximum number of collected links per collected resources	Maximum number of unique resolved links per selected resources
heiwww	13.301	2.783	339	145
liawww	12.309	1.703	95	22
iamwww	30.735	1.497	11625	623
isoe	11.772	2.136	32	19
isbiel	26.004	3.480	2191	365
tecfa	20.274	3.502	1492	309
tecfamoo	3.536	2.612	41	28
depth	21.099	1.987	1507	196
breadth	25.863	2.926	5965	115
ch	25.589	2.264	7834	207

Dataset	Mean number of collected links per collected resources	Mean number of unique resolved links per selected resources	Maximum number of collected links per collected resources	Maximum number of unique resolved links per selected resources
ops	28.495	1.619	1816	171
unine	10.478	1.801	228	18

Dataset	Linkage density of the collected resources	Resolved linkage density of the collected resources	Resolved linkage density of the selected resources	Mean distance among resources in the graph G	Maximum distance among resource in the graph G
heiwww	0.0665	0.0221	0.0139	3.8861	9
liawww	0.1598	0.0638	0.0221	5.4786	12
iamwww	0.0011	0.0041	0.0029	2.6324	5
isoe	0.5475	0.1183	0.0993	3.1807	7
isbiel	0.0529	0.0299	0.0070	3.3457	8
tecfa	0.0080	0.0264	0.0070	3.8161	7
tecfamoo	0.0067	0.0064	0.0049	7.9270	19
depth	0.0264	0.0090	0.0039	23.1570	50
breadth	0.0051	0.0006	0.0585	3,6252	4
ch	0.2939	0.0028	0.0045	3.7937	5
ops	0.0056	0.0004	0.0032	4.2860	7
unine	0.1309	0.0338	0.0225	4.9310	10