Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)
http://dare.uva.nl/document/196833

File ID	196833
Filename	Chapter 2 Information Retrieval Systems

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	Dissertation
Title	Giving Eyes to ICT! Or How does a computer recognize a cow?
Author	B.A.M. Schouten
Faculty	Faculty of Science
Year	2001

FULL BIBLIOGRAPHIC DETAILS:
http://dare.uva.nl/record/96150

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other then for strictly personal, individual use.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)

Chapter 2

Information Retrieval Systems

A text document can be well described by keywords. In the extreme case as many keywords as words in the document are needed to retrieve the document. Together the keywords constitute a description of the content. Visual content is more like poetry. It can have several layered meanings at the same time and is strongly dependent on the appreciation of the user. For this reason it is much more difficult to describe its content. Historically, document retrieval is done by keyword search, where no distinction is made to the nature of the document whether visual or textual.

In the last 10 years, with advancing information technology, image analysis and pattern recognition provided tools to access visual information according to human perception. Visual content descriptors used in contemporary multimedia applications describe the content in low level terms, like the color distribution of an image. These descriptors can be used to define a similarity between documents, such that they can be searched for in database systems.

2.1 Relational Database

In 1978 the ANSI/X3/SPARC Study Group [33] proposed to add an extra dimension to the way *database management systems* (DBMS) function. Until then DBMS envisioned a two level organization: the data as seen by the system (*internal level*) and the data as seen by the user (*external level*). The internal level takes care of the storage of the data. The external level consists of user views that enable the user to communicate with the database and to retrieve the data. A newly created third level, called the *conceptual level*, represents the entities, properties and relationships of the documents stored in the database. In this way relations are used to retrieve information from the database. The introduction of the third level was the starting point of the *relational database*.

2.2 First Generation Visual Information Retrieval Systems

In first generation visual information retrieval (VIR) systems, in order to query a visual or textual document one is invited to provide a keyword which (supposedly) labels the document. This keyword is "processed" in the database and a perfect match at keyword level is aimed for. An expert off-line has supplied the keywords, for a schematic overview see Figure 2.1. The views presented at the external level consist of search engines addressed by traditional query languages like *SQL*.

Standard contemporary systems are able to relate and associate between keywords, in this way these systems build upon the principles of the relational database, see Section 2.1. Semantically more complex queries can be made by Boolean expressions. If the relationship between the keywords is provided by natural language processing, semantically more intelligent queries can be made. Even more advanced systems provide the user with the possibility to browse the database. To search a large variety of multimedia data efficiently, browsing is a desired functionality; it enables the user to have an overview of the data. For browsing, visualization of the data space plays a crucial role [21, 14] (Figure 2.2). Click A.2

The limitations of the first generation systems are:

- The conceptual level is not created by the user but manually by an "expert": the descriptions of the content are made by an annotator. The user may find these choices arbitrary, even inappropriate.
- Content descriptors are created manually. Automatic or semi-automatic content description is desirable.
- Some content can hardly be described in any language e.g. textures or a painting by Picasso.
- The relations between the keywords are also described in the text domain. Concepts like similarity are hard to express in this domain.

As an illustration of these limitations, please have a look at the result of a search action using the word *koe* [1] within the popular search engine *Altavista*. Click B.3

2.3 Second Generation Visual Information Retrieval Systems

Multimedia demands a more sophisticated way of storing and accessing the objects within the database. At the internal level one sees a tendency to use more advanced databases able

[1] Meaning cow in Dutch.

to store complex data consisting of video, images, sound etc. The inclusion of multimedia in a database has a profound impact on its design, features and functions. If a database is only storing multimedia for delivery, like in first generation VIR systems, then a multimedia capable file server coupled with the ability to store pointers, filenames or object identifiers in the database is sufficient.

Utilizing deeper semantic knowledge about the media such as the ability to index, search and relate information, is truly a function which adds value to a database system. It needs to provide support for a range of existing and future data-types, and needs to include support for both temporal and spatial modeling such that data abstraction is maximized [15]. The Monet database at CWI is an example:

`http://dbs.cwi.nl:8080/cwwwi/owa/cwwwi.print_projects?ID=41.`

In visual information systems, feature extractors yield descriptions of the visual content. Data independence is important to easily update and change the information contained in a database. The *MPEG* standards provide an example. Click A.3

2.3.1 Visual Content

Metadata describing the visual content of images can be subdivided into [12]:

- **Content independent.**
  Data like author, date, size.
- **Content dependent.**
  - **Perceptual.** Perceptual facts within a video or image. Typically describing aspects like color, texture, shape etc within an image or video.
  - **Semantic.** Abstractions like objects in an image or a scene in a movie.
- **Psychological.**
  Data describing human values and emotions.

Improvements in the second generation retrieval systems are made at the perceptual level. Access to visual information is not merely by keywords but also by using objective measurements of the visual content. Firstly, we attach *features* to images. Usually a feature is some vector in $\mathbb{R}^n$ and the result of the analysis of pixel distributions and numerical discretization of perceptual properties. Secondly, these features are precompiled for all images in the database. In a broad sense features may include keyword based features as well as visual features. For a schematic overview see Figure 2.3.

The user has several methods to provide input for the system, e.g. by sketching or select- Click A.4

ing from image samples. The user selects which features are relevant to the image retrieval process and indicates which similarity measure should be used. The system compares the features of all the images with the query, according to the selected similarity measure and returns an index of images *similar* to the query image provided by the user.

Limitations of the system are:

- **Subjectivity of the features.**
  - Mathematically defined features may not have a clear perceptual meaning or interpretation.
  - Each feature describes just one aspect.
  - Features are driven by availability and not by necessity.
  - High level aspects of an image like objects can not sufficiently be described by (combinations of) low level features like color or texture. It is impossible to describe a beautiful painting of Picasso merely by the colors used.
- **Subjectivity of the user.**
  The content of an image is subjective and strongly dependent on culture, personal taste and different opinions the user can have at different times.
- **Subjectivity of the image.**
  An image may consist of many scenes, see Figure 2.4.

2.3.2 Examples

The goal is to recognize images by their content. Historically these images were manually described by keywords. In early examples of content based image retrieval like Photobook [23] and QBIC[22], one can select an image in the database and ask the system to retrieve images that look similar according to some feature like color, texture, graphical design etc. The user could indicate how strongly the feature should be taken into account when searching for similar images. A sketch interface could be used to indicate and locate different objects in an image.
Similarity is defined on the actual content of the image. Content is derived using image analysis tools. Photobook was one of the first systems for browsing and retrieval of still images by image content. Several subsystems according to the class of images emerged. One of the most well known subsystems is the one dedicated to face-similarity [23].

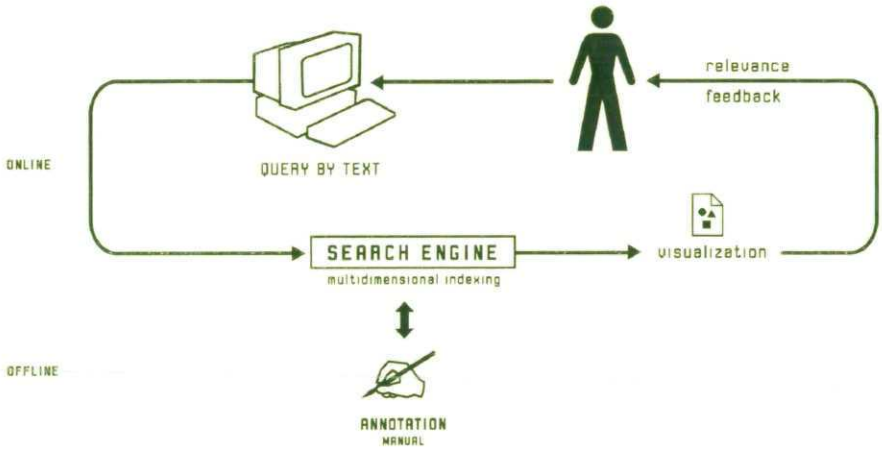

Figure 2.1: Schematic overview of first generation visual information retrieval system.

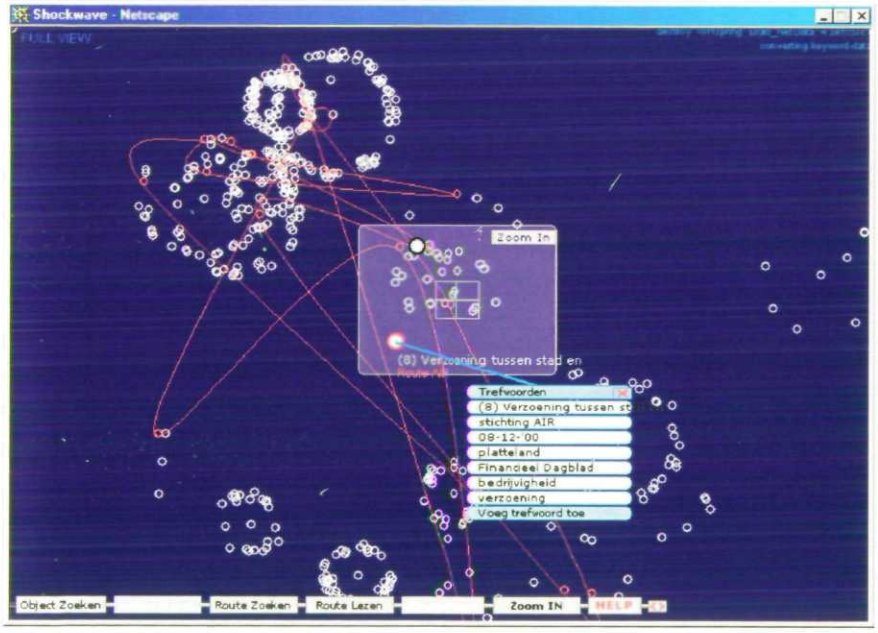

Figure 2.2: A datacloud showing an overview of the data in the database.

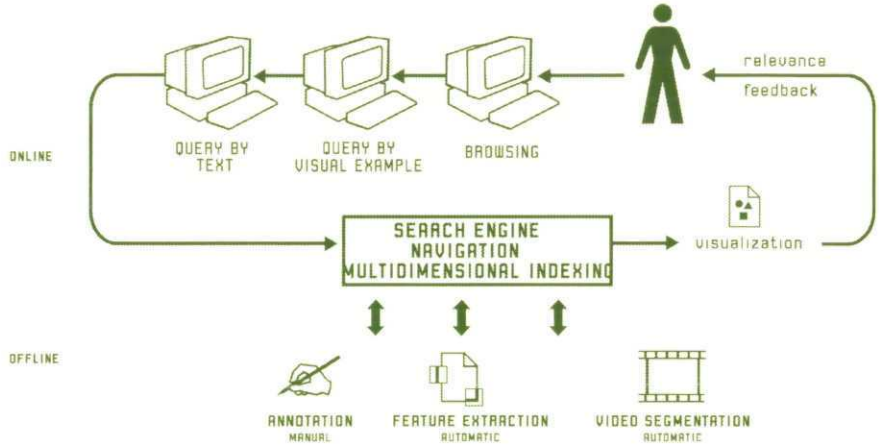

Figure 2.3: Schematic overview of second generation visual information retrieval system.

Figure 2.4: HRH Prince William of Orange milking a cow. The image shown can be appreciated for several reasons: The beautifully made *papier-mâché* cow. The infant. His Royal Highness Prince William of Orange.