**Module 6**

**User Interface and Applications**

<https://en.wikipedia.org/wiki/Information_retrieval>

# **User interface for information retrieval system:**

An improved information retrieval system user interface for retrieving information from a plurality of sources and for storing information source descriptions in a knowledge base. The user interface includes a hypertext browser and a knowledge base browser/editor. The hypertext browser allows a user to browse an unstructured information space through the use of interactive hypertext links. The knowledge base browser/editor displays a directed graph representing a generalization taxonomy of the knowledge base, with the nodes representing concepts and edges representing relationships between concepts. The system allows users to store information source descriptions in the knowledge base via graphical pointing means. By dragging an iconic representation of an information source from the hypertext browser to a node in the directed graph, the system will store an information source description object in the knowledge base. The knowledge base browser/editor is also used to browse the information source descriptions previously stored in the knowledge base. The result of such browsing is an interactive list of information source descriptions which may be used to retrieve documents into the hypertext browser. The system also allows for querying a structured information source and using query results to focus the hypertext browser on the most relevant unstructured data sources.

**1.Video information retrieval:**

A **video search engine** is a web-based [search engine](https://en.wikipedia.org/wiki/Search_engine) which [crawls](https://en.wikipedia.org/wiki/Web_crawler) the web for [video](https://en.wikipedia.org/wiki/Video) content. Some video search engines parse externally hosted content while others allow content to be uploaded and hosted on their own servers. Some engines also allow users to search by video format type and by length of the clip. The video search results are usually accompanied by a [thumbnail](https://en.wikipedia.org/wiki/Thumbnail) view of the video.

Video search engines are computer programs designed to find videos stored on digital devices, either through Internet servers or in storage units from the same computer. These searches can be made through audiovisual [indexing](https://en.wikipedia.org/wiki/Search_engine_indexing), which can extract information from audiovisual material and record it as metadata, which will be tracked by search engines.

## **1.1 Utility**

The main use of these search engines is the increasing creation of audiovisual content and the need to manage it properly. The digitization of audiovisual archives and the establishment of the Internet, has led to large quantities of video files stored in big databases, whose recovery can be very difficult because of the huge volumes of data and the existence of a semantic gap.

## **1.2 Search criterion**

The search criterion used by each search engine depends on its nature and purpose of the searches.

### Metadata

Metadata is information about facts. It could be information about who is the author of the video, creation date, duration, and all the information that could be extracted and included in the same files. Internet is often used in a language called XML to encode metadata, which works very well through the web and is readable by people. Thus, through this information contained in these files is the easiest way to find data of interest to us.

In the videos there are two types of metadata, that we can integrate in the video code itself and external metadata from the page where the video is. In both cases we optimize them to make them ideal when indexed.

#### Internal metadata

All video formats incorporate their own metadata. The title, description, coding quality or transcription of the content are possible. To review these data exist programs like FLV MetaData Injector, Sorenson Squeeze or Castfire. Each one has some utilities and special specifications.

Converting from one format to another can lose much of this data, so check that the new format information is correct. It is therefore advisable to have the video in multiple formats, so all search robots will be able to find and index it.

#### External metadata

In most cases the same mechanisms must be applied as in the positioning of an image or text content.

##### Title and description

They are the most important factors when positioning a video, because they contain most of the necessary information. The titles have to be clearly descriptive and should remove every word or phrase that is not useful.

##### Filename

It should be descriptive, including keywords that describe the video with no need to see their title or description. Ideally, separate the words by dashes "-".

##### Tags

On the page where the video is, it should be a list of keywords linked to the microformat "rel-tag". These words will be used by search engines as a basis for organizing information.

##### Transcription and subtitles

Although not completely standard, there are two formats that store information in a temporal component that is specified, one for subtitles and another for transcripts, which can also be used for subtitles. The formats are SRT or SUB for subtitles and TTXT for transcripts.

### Speech recognition

[Speech recognition](https://en.wikipedia.org/wiki/Speech_recognition) consists of a transcript of the speech of the audio track of the videos, creating a text file. In this way and with the help of a phrase extractor can easily search if the video content is of interest. Some search engines apart from using speech recognition to search for videos, also use it to find the specific point of a multimedia file in which a specific word or phrase is located and so go directly to this point. Gaudi (Google Audio Indexing), a project developed by [Google Labs](https://en.wikipedia.org/wiki/Google_Labs), uses voice recognition technology to locate the exact moment that one or more words have been spoken within an audio, allowing the user to go directly to exact moment that the words were spoken. If the search query matches some videos from YouTube, the positions are indicated by yellow markers, and must pass the mouse over to read the transcribed text.

#### Speaker recognition

In addition to transcription, analysis can detect different speakers and sometime attribute the speech to an identified name for the speaker.

### Text recognition

The text recognition can be very useful to recognize characters in the videos through "chyrons". As with speech recognizers, there are search engines that allow (through character recognition) to play a video from a particular point.

TalkMiner, an example of search of specific fragments from videos by text recognition, analyzes each video once per second looking for identifier signs of a slide, such as its shape and static nature, captures the image of the slide and uses [Optical Character Recognition](https://en.wikipedia.org/wiki/Optical_character_recognition) (OCR) to detect the words on the slides. Then, these words are indexed in the [search engine](https://en.wikipedia.org/wiki/Search_engine) of TalkMiner, which currently offers to users more than 20,000 videos from institutions such as Stanford University, the University of California at Berkeley, and TED.

### Frame analysis

Through the [visual descriptors](https://en.wikipedia.org/wiki/Visual_descriptors) we can analyze the frames of a video and extract information that can be scored as metadata. Descriptions are generated automatically and can describe different aspects of the frames, such as color, texture, shape, motion, and the situation.

#### Chaptering

The video analysis can lead to automatic chaptering, using technics such as change of camera angle, identification of audio jingles. By knowing the typical structure of a video document, it is possible to identify starting and ending credits, content parts and beginning and ending of advertising breaks.

## **1.3 Ranking criterion**

The usefulness of a search engine depends on the [relevance](https://en.wikipedia.org/wiki/Relevance_(information_retrieval)) of the result set returned. While there may be millions of videos that include a particular word or phrase, some videos may be more relevant, popular or have more authority than others. This arrangement has a lot to do with search engine optimization.

Most search engines use different methods to classify the results and provide the best video in the first results. However, most programs allow sorting the results by several criteria.

### Order by relevance

This criterion is more ambiguous and less objective, but sometimes it is the closest to what we want; depends entirely on the searcher and the algorithm that the owner has chosen. That's why it has always been discussed and now that search results are so ingrained into our society it has been discussed even more. This type of management often depends on the number of times that the searched word comes out, the number of viewings of this, the number of pages that link to this content and ratings given by users who have seen it.[[1]](https://en.wikipedia.org/wiki/Video_search_engine#cite_note-1)

### Order by date of upload

This is a criterion based totally on timeline. Results can be sorted according to their seniority in the repository.

### Order by number of views

It can give us an idea of the popularity of each video.

### Order by length

This is the length of the video and can give a taste of which video it is.

### Order by user rating

It is common practice in repositories let the users rate the videos, so that a content of quality and relevance will have a high rank on the list of results gaining visibility. This practice is closely related to virtual communities.

**2. Image information retrieval**

An image retrieval system is a computer system used for browsing, searching and retrieving images from a large [database](https://en.wikipedia.org/wiki/Database) of digital images. Most traditional and common methods of image retrieval utilize some method of adding [metadata](https://en.wikipedia.org/wiki/Metadata) such as [captioning](https://en.wikipedia.org/wiki/Photo_caption), [keywords](https://en.wikipedia.org/wiki/Keyword_(internet_search)), title or descriptions to the images so that retrieval can be performed over the annotation words. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social [web applications](https://en.wikipedia.org/wiki/Web_application) and the [semantic web](https://en.wikipedia.org/wiki/Semantic_web) have inspired the development of several web-based image annotation tools.

## **2.1 Search methods**

**Image search** is a specialized data search used to find images. To search for images, a user may provide query terms such as keyword, image file/link, or click on some image, and the system will return images "similar" to the query. The similarity used for search criteria could be meta tags, color distribution in images, region/shape attributes, etc.

* [Image meta search](https://en.wikipedia.org/wiki/Image_meta_search) - search of images based on associated metadata such as keywords, text, etc.
* [Content-based image retrieval](https://en.wikipedia.org/wiki/Content-based_image_retrieval) (CBIR) – the application of [computer vision](https://en.wikipedia.org/wiki/Computer_vision) to the image retrieval. CBIR aims at avoiding the use of textual descriptions and instead retrieves images based on similarities in their contents (textures, colors, shapes etc.) to a user-supplied query image or user-specified image features.
  + [List of CBIR Engines](https://en.wikipedia.org/wiki/List_of_CBIR_Engines) - list of engines which search for images based image visual content such as color, texture, shape/object, etc.

*Further information:*[*Visual search engine*](https://en.wikipedia.org/wiki/Visual_search_engine)*and*[*Reverse image search*](https://en.wikipedia.org/wiki/Reverse_image_search)

* [Image collection exploration](https://en.wikipedia.org/wiki/Image_collection_exploration) - search of images based on the use of novel exploration paradigms.[[3]](https://en.wikipedia.org/wiki/Image_retrieval#cite_note-3)

## Data scope[[edit](https://en.wikipedia.org/w/index.php?title=Image_retrieval&action=edit&section=2)]

It is crucial to understand the scope and nature of image data in order to determine the complexity of image search system design. The design is also largely influenced by factors such as the diversity of user-base and expected user traffic for a search system. Along this dimension, search data can be classified into the following categories:

* *Archives* - usually contain large volumes of structured or semi-structured homogeneous data pertaining to specific topics.
* *Domain-Specific Collection* - this is a homogeneous collection providing access to controlled users with very specific objectives. Examples of such a collection are biomedical and satellite image databases.
* *Enterprise Collection* - a heterogeneous collection of images that is accessible to users within an organization's intranet. Pictures may be stored in many different locations.
* *Personal Collection* - usually consists of a largely homogeneous collection and is generally small in size, accessible primarily to its owner, and usually stored on a local storage media.
* *Web* - World Wide Web images are accessible to everyone with an Internet connection. These image collections are semi-structured, non-homogeneous and massive in volume, and are usually stored in large disk arrays.

**3. 3D retrieval**

A 3D Content Retrieval system is a computer system for browsing, searching and retrieving three dimensional digital contents (e.g.: [Computer-aided design](https://en.wikipedia.org/wiki/Computer-aided_design), molecular biology models, and cultural heritage 3D scenes, etc.) from a large database of digital images. The most original way of doing 3D content retrieval uses methods to add description text to 3D content files such as the content file name, link text, and the web page title so that related 3D content can be found through text retrieval. Because of the inefficiency of manually annotating 3D files, researchers have investigated ways to automate the [annotation](https://en.wikipedia.org/wiki/Automatic_image_annotation) process and provide a unified standard to create text descriptions for 3D contents. Moreover, the increase in 3D content has demanded and inspired more advanced ways to retrieve 3D information. Thus, shape matching methods for 3D content retrieval have become popular. Shape matching retrieval is based on techniques that compare and contrast similarities between 3D models.

## **3.1 3D retrieval methods**

**Derive a high level description (e.g.: a skeleton) and then find matching results**

This method describes 3D models by using a skeleton. The skeleton encodes the geometric and topological information in the form of a skeletal graph and uses [graph matching](https://en.wikipedia.org/wiki/Graph_matching) techniques to match the skeletons and compare them.[[1]](https://en.wikipedia.org/wiki/3D_Content_Retrieval#cite_note-1) However, this method requires a 2-manifold input model, and it is very sensitive to noise and details. Many of the existing 3D models are created for visualization purposes, while missing the input quality standard for the skeleton method. The skeleton 3D retrieval method needs more time and effort before it can be used widely.

**Compute a feature vector based on statistics**

Unlike Skeleton modeling, which requires a high quality standard for the input source, statistical methods do not put restriction on the validity of an input source. Shape [histograms](https://en.wikipedia.org/wiki/Histogram), feature vectors composed of global geo-metic properties such as circularity and eccentricity, and feature vectors created using frequency decomposition of spherical functions are common examples of using statistical methods to describe 3D information.[[2]](https://en.wikipedia.org/wiki/3D_Content_Retrieval#cite_note-2)

**2D projection method**

Some approaches use 2D projections of a 3D model, justified by the assumption that if two objects are similar in 3D, then they should have similar 2D projections in many directions. [Prototypical](https://en.wikipedia.org/wiki/Prototype) Views[[3]](https://en.wikipedia.org/wiki/3D_Content_Retrieval#cite_note-3) and [Light field](https://en.wikipedia.org/wiki/Light_field) description[[4]](https://en.wikipedia.org/wiki/3D_Content_Retrieval#cite_note-4) are good examples of 2D projection methods.

## **3.2 Challenges**

**Challenges associated with 3D shape-based similarity queries**

With the skeleton modeling 3D retrieval method, figuring out an efficient way to index 3D shape descriptors is very challenging because 3D shape indexing has very strict criteria. The 3D models must be quick to compute, concise to store, easy to index, invariant under similarity transformations, insensitive to noise and small extra features, robust to arbitrary topological degeneracies, and discriminating of shape differences at many scales.

**3D search and retrieval with multimodal support challenges**

In order to make the 3D search interface simple enough for novice users who know little on 3D retrieval input source requirements, a [multimodal](https://en.wikipedia.org/wiki/Multimodal_interaction) retrieval system, which can take various types of input sources and provide robust query results, is necessary. So far, only a few approaches have been proposed. In Funkhouser et al. (2003),[[6]](https://en.wikipedia.org/wiki/3D_Content_Retrieval#cite_note-6) the proposed “Princeton 3D search engine” supports 2D sketches, 3D sketches, 3D models and text as queries. In Chen et al. (2003),[[7]](https://en.wikipedia.org/wiki/3D_Content_Retrieval#cite_note-7) he designed a 3D retrieval system that intakes 2D sketches and retrieves for 3D objects. Recently, Ansary et al. (2007)[[8]](https://en.wikipedia.org/wiki/3D_Content_Retrieval#cite_note-8) proposed a 3D retrieval framework using 2D photographic images, sketches, and 3D models.

**4. Audio and Music retrieval**

Music information retrieval (MIR) is the interdisciplinary science of retrieving [information](https://en.wikipedia.org/wiki/Information) from [music](https://en.wikipedia.org/wiki/Music). Those involved in MIR may have a background in academic [musicology](https://en.wikipedia.org/wiki/Musicology), [psychoacoustics](https://en.wikipedia.org/wiki/Psychoacoustics), [psychology](https://en.wikipedia.org/wiki/Psychology), [signal processing](https://en.wikipedia.org/wiki/Signal_processing), [informatics](https://en.wikipedia.org/wiki/Informatics), [machine learning](https://en.wikipedia.org/wiki/Machine_learning), [optical music recognition](https://en.wikipedia.org/wiki/Optical_music_recognition), [computational intelligence](https://en.wikipedia.org/wiki/Computational_intelligence) or some combination of these.

## **4.1 Applications**

MIR is being used by businesses and academics to categorize, manipulate and even create music.

### Music classification

One of the classical MIR research topic is genre classification, which is categorizing music items into one of pre-defined genres such as [classical](https://en.wikipedia.org/wiki/Classical_music), [jazz](https://en.wikipedia.org/wiki/Jazz), [rock](https://en.wikipedia.org/wiki/Rock_music), etc. [Mood classification](https://en.wikipedia.org/wiki/Emotion_classification), artist classification, instrument identification, and music tagging are also popular topics.

### Recommender systems

Several [recommender systems](https://en.wikipedia.org/wiki/Recommender_systems) for music already exist, but surprisingly few are based upon MIR techniques, instead making use of similarity between users or laborious data compilation. [Pandora](https://en.wikipedia.org/wiki/Pandora_Radio), for example, uses experts to tag the music with particular qualities such as "female singer" or "strong bassline". Many other systems find users whose listening history is similar and suggests unheard music to the users from their respective collections. MIR techniques for [similarity in music](https://en.wikipedia.org/wiki/Musical_similarity) are now beginning to form part of such systems.

### Music source separation and instrument recognition

Music source separation is about separating original signals from a mixture [audio signal](https://en.wikipedia.org/wiki/Audio_signal). Instrument recognition is about identifying the instruments involved in music. Various MIR systems have been developed that can separate music into its component tracks without access to the master copy. In this way e.g. karaoke tracks can be created from normal music tracks, though the process is not yet perfect owing to vocals occupying some of the same [frequency](https://en.wikipedia.org/wiki/Frequency) space as the other instruments.

### Automatic music transcription

Automatic [music transcription](https://en.wikipedia.org/wiki/Transcription_(music)) is the process of converting an audio recording into symbolic notation, such as a score or a [MIDI file](https://en.wikipedia.org/wiki/MIDI_file#File_formats).[[1]](https://en.wikipedia.org/wiki/Music_information_retrieval#cite_note-1) This process involves several audio analysis tasks, which may include multi-pitch detection, [onset detection](https://en.wikipedia.org/wiki/Onset_detection#Onset_detection), duration estimation, instrument identification, and the extraction of [harmonic](https://en.wikipedia.org/wiki/Harmonic), [rhythmic](https://en.wikipedia.org/wiki/Rhythm) or [melodic](https://en.wikipedia.org/wiki/Melody) information. This task becomes more difficult with greater numbers of instruments and a greater [polyphony level](https://en.wikipedia.org/wiki/Polyphony_and_monophony_in_instruments).

### Music generation

The [automatic generation of music](https://en.wikipedia.org/wiki/Automatic_generation_of_music) is a goal held by many MIR researchers. Attempts have been made with limited success in terms of human appreciation of the results.

## **4.2 Methods used**

### Data source

[Scores](https://en.wikipedia.org/wiki/Sheet_music) give a clear and logical description of music from which to work, but access to [sheet music](https://en.wikipedia.org/wiki/Sheet_music), whether digital or otherwise, is often impractical. [MIDI](https://en.wikipedia.org/wiki/MIDI) music has also been used for similar reasons, but some data is lost in the conversion to MIDI from any other format, unless the music was written with the MIDI standards in mind, which is rare. [Digital audio formats](https://en.wikipedia.org/wiki/Audio_file_format) such as [WAV](https://en.wikipedia.org/wiki/WAV), [mp3](https://en.wikipedia.org/wiki/Mp3), and [ogg](https://en.wikipedia.org/wiki/Ogg) are used when the audio itself is part of the analysis. Lossy formats such as mp3 and ogg work well with the human ear but may be missing crucial data for study. Additionally some encodings create artifacts which could be misleading to any automatic analyser. Despite this the ubiquity of the mp3 has meant much research in the field involves these as the source material. Increasingly, [metadata](https://en.wikipedia.org/wiki/Metadata) mined from the web is incorporated in MIR for a more rounded understanding of the music within its cultural context, and this recently consists of analysis of [social tags](https://en.wikipedia.org/wiki/Social_tagging) for music.

### Feature representation

Analysis can often require some summarising,[[2]](https://en.wikipedia.org/wiki/Music_information_retrieval#cite_note-2) and for music (as with many other forms of data) this is achieved by [feature extraction](https://en.wikipedia.org/wiki/Feature_extraction), especially when the [audio content](https://en.wikipedia.org/wiki/Computer_audition) itself is analysed and [machine learning](https://en.wikipedia.org/wiki/Machine_learning) is to be applied. The purpose is to reduce the sheer quantity of data down to a manageable set of values so that learning can be performed within a reasonable time-frame. One common feature extracted is the [Mel-Frequency Cepstral Coefficient](https://en.wikipedia.org/wiki/Mel-frequency_cepstral_coefficient) (MFCC) which is a measure of the [timbre](https://en.wikipedia.org/wiki/Timbre) of a [piece of music](https://en.wikipedia.org/wiki/Musical_composition). Other features may be employed to represent the [key](https://en.wikipedia.org/wiki/Tonality#Computational_methods_to_determine_the_key), [chords](https://en.wikipedia.org/wiki/Chord_(music)), [harmonies](https://en.wikipedia.org/wiki/Harmony), [melody](https://en.wikipedia.org/wiki/Melody), main [pitch](https://en.wikipedia.org/wiki/Pitch_(music)), [beats per minute](https://en.wikipedia.org/wiki/Beats_per_minute) or rhythm in the piece. There are a number of available audio feature extraction tools[[3]](https://en.wikipedia.org/wiki/Music_information_retrieval#cite_note-3) [Available here](https://www.ntnu.edu/documents/1001201110/1266017954/DAFx-15_submission_43_v2.pdf)

### Statistics and machine learning

* Computational methods for classification, clustering, and modelling — musical feature extraction for mono- and [polyphonic](https://en.wikipedia.org/wiki/Polyphonic) music, similarity and [pattern matching](https://en.wikipedia.org/wiki/Pattern_matching), retrieval
* [Formal methods](https://en.wikipedia.org/wiki/Formal_methods) and [databases](https://en.wikipedia.org/wiki/Database) — applications of automated [music identification](https://en.wikipedia.org/wiki/Music_identification) and recognition, such as [score following](https://en.wikipedia.org/wiki/Score_following), automatic accompaniment, routing and filtering for music and music queries, query languages, standards and other metadata or protocols for music information handling and [retrieval](https://en.wikipedia.org/wiki/Information_retrieval), [multi-agent systems](https://en.wikipedia.org/wiki/Multi-agent_system), distributed search)
* Software for music information retrieval — [Semantic Web](https://en.wikipedia.org/wiki/Semantic_Web) and musical digital objects, [intelligent agents](https://en.wikipedia.org/wiki/Intelligent_agent), [collaborative software](https://en.wikipedia.org/wiki/Collaborative_software), web-based search and [semantic retrieval](https://en.wikipedia.org/w/index.php?title=Semantic_retrieval&action=edit&redlink=1), [query by humming](https://en.wikipedia.org/wiki/Query_by_humming) / [Search by sound](https://en.wikipedia.org/wiki/Search_by_sound), [acoustic fingerprinting](https://en.wikipedia.org/wiki/Acoustic_fingerprinting)
* Music analysis and knowledge representation — [automatic summarization](https://en.wikipedia.org/wiki/Automatic_summarization), citing, excerpting, downgrading, transformation, formal models of music, digital scores and representations, music indexing and [metadata](https://en.wikipedia.org/wiki/Metadata).