TECHNICAL REPORT

ISO/IEC TR 24800-1

First edition
2007-12-15

Information technology — JPSearch —

Part 1: System framework and components

Technologies de l'information — JPSearch —

Partie 1: Cadre système et composants

Reference number
ISO/IEC TR 24800-1:2007(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword iv

Introduction v

1 Scope 1
1.1 Interoperable Image Search and Retrieval 1
1.2 Motivation 1
1.3 Outline of the Technical Report 2

2 Terms, Definitions and Abbreviated Terms 3
2.1 Terms and definitions 3
2.2 Symbols and abbreviated terms 4

3 Background and motivation of a user-centric approach to image search 4
3.1 Traditional Models of Image Search and Retrieval 4
3.2 The user as part of the retrieval system 5
3.3 User task and user evaluation 6
3.4 Meaning also comes from outside 8

4 Use Cases 11
4.1 Introduction 11
4.2 Searching images in stock photo collections for usage in magazines 11
4.3 Searching for and publishing authoritative themed sub-collections of images 11
4.4 Mobile Tourist Information 11
4.5 Surveillance Search from Desktop to Mobile Device with Alerts 11
4.6 Ad hoc search without time-consuming housekeeping tasks 11
4.7 Rights clearance to publish a compliant business document 11
4.8 Tracking an object creation process using a temporal series of photos 11
4.9 Finding illegal or unauthorized use of images 11
4.10 Finding the best shots or filtering out of the worse shots 12
4.11 Context searching without human annotation 12
4.12 Image search based on image quality 12
4.13 Image search with deduplication 12
4.14 Matching images between collections for synchronization 12
4.15 Social metadata updating and sharing of images for searching 12
4.16 Image search in the medical domain 12
4.17 Servants image searchers 12
4.18 Open federated repositories 13

5 Image search and management process 13
5.1 Introduction 13
5.2 Metadata flow 13
5.3 Query process flow 15

6 JPSearch Architecture 18
6.1 Overview of the architecture 18
6.2 Parts to be standardized 20
6.3 Architecture justification with use cases 20

7 Organization of the JPSearch specification 21
7.1 Overall Structure of JPSearch 21
7.2 Part 2: Schema and ontology registration and identification 21
7.3 Part 3: JPSearch query format 22
7.4 Part 4: Metadata embedded in image data (JPEG and JPEG 2000) file format 22
7.5 Part 5: Data interchange format between image repositories 22

Annex A (informative) Use Cases 24

Bibliography 37

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

In exceptional circumstances, the joint technical committee may propose the publication of a Technical Report of one of the following types:

— type 1, when the required support cannot be obtained for the publication of an International Standard, despite repeated efforts;

— type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;

— type 3, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example).

Technical Reports of types 1 and 2 are subject to review within three years of publication, to decide whether they can be transformed into International Standards. Technical Reports of type 3 do not necessarily have to be reviewed until the data they provide are considered to be no longer valid or useful.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR 24800-1, which is a Technical Report of type [3], was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

ISO/IEC TR 24800 consists of the following parts, under the general title *Information technology — JPSearch*:

— *Part 1: System framework and components*

Introduction

JPSearch aims to provide a standard for interoperability for still image search and retrieval systems. There are many systems that provide image search and retrieval functionality on computer desktops, on the World Wide Web (i.e. websearch), on imaging devices, and in other consumer and professional applications. Existing systems are implemented in a way that tightly couples many components of the search process. JPSearch provides an abstract framework search architecture that decouples the components of image search and provides a standard interface between these components.

Aligning image search system design to this standard framework facilitates the use and reuse of metadata; the use and reuse of profiles and ontologies to provide a common context for searching; the provision of a common query language to search easily across multiple repositories with the same search semantics; allows image repositories to be independent of particular system implementations; and for users to move easily or upgrade their image management applications or to move to a different device or upgrade to a new computer.

Information technology — JPSearch —

Part 1:
System framework and components

1 Scope

1.1 Interoperable Image Search and Retrieval

This Technical Report specifies two things. The first is a framework for interoperability for still image search and retrieval. The second identifies an architecture and the components in this framework, the linkages between components, and which of these components and links are to be standardized in JPSearch.

The image search and retrieval framework will be determined by real use cases (tasks) and will leverage on lessons learnt in the long history of text retrieval where, for example, different users issuing the same query may be looking for (very) different results. This is important because it means that the framework must be general enough to support many possible approaches to image retrieval, e.g., from using only low-level image features, to text annotations, to community input, or a mixture of such approaches.

From the framework and components, and the linkages and flow of data between them, the parts of JPSearch that need to be standardized can be determined.

1.2 Motivation

There are many applications that provide image search and retrieval functionality on computer desktops, on the World Wide Web (i.e., websearch), on imaging devices, and in other consumer and professional applications. These implementations are characterized by significant limitations, including:

- **Lack of the ability to reuse metadata**

  The biggest problem in still image management is consistent and complete user or system annotation (in whatever form) of images. A user makes a heavy investment if and when they annotate an image or a collection of images. For example, a user adopts System A for storing and managing still images. The user discovers System B, which provides improved and desirable functionality, but is effectively prevented from switching to System B because the metadata in System A cannot be easily (or at all) used in System B. *In this example, users are impeded in using the applications or systems that best suit their needs; and system providers are unable to compete freely with their products*.

  This problem generalizes in community based image sharing systems, where multiple users may annotate the shared images. In most cases, however, an image has a single owner and there is a need for the ability to merge community metadata back into the owner's image management system. *This ability would help overcome the difficult problem of manual still image annotation*.

- **Lack of a common query format and search semantics**

  There is a trend towards shared image repositories. These could be on the web, but there are also systems that publish user repositories residing on their local (e.g., home) machines for (normally access controlled) public viewing and annotation. *As the number and size of such repositories increase (a monotonic increasing trend), search becomes an essential function for users to navigate shared repositories*.

Unfortunately, the various systems providing image search, whether on the desktop or on the web, do not provide a common way of specifying a search. This is not the same as having a common user-interface since the look and feel is up to a system provider to provide and for the user to like or not like. The problem is that a query such as “white car” may be interpreted as a Boolean “white” AND “car” or “white” OR “car”, or “white car” as a phrase, etc., and the interpretation may be different when the search is done against the image data or against text metadata, or against other metadata. *Users are confused because different systems return different results for the same query. System providers need a reference standard to remove ambiguity and make searching over shared repositories consistent*.

- **Lack of a common format for handling context in searching**

  A large adult describing a 5-foot tall man may use the word “short”. A small child looking at the same person may say “tall”. This does not mean that the person is both “short” and “tall” at the same time; rather it is the context that has changed. Similarly, when a doctor does a query using the term “skin cancer”, he or she probably expects a very different set of images from when a patient searches with the same query term. *Searching for images always takes place in a context. This context may be implicit or explicit*.

  Some systems allow the user to specify a context and there are other systems that automatically imply a user’s context. There is no way for the context in one search system to be used in a different search system. *A common format for handling context allows a user to carry their context with them to different search engines. It also allows the context to be owned by the user and not by the system, i.e., it protects the user’s privacy.*

These are just three examples of where still image search systems can benefit tremendously from interoperability. Other examples include how metadata can be created, evolved and stored, and also how image collections can have metadata different from and augmenting the metadata of a single image.

Existing systems are implemented in a way that tightly couples many components of the search process. JPSearch provides an abstract framework search architecture that allows an alignment of system design to a standard framework. Among other things, this alignment facilitates the use and reuse of metadata, the use and reuse of ontologies to provide a common language for contexts, the provision of a common query language, provide standardized interface to system components, and the ability to provide still image search and retrieval functionality across multiple repositories.

1.3 Outline of the Technical Report

There will be 7 clauses to this report. They are arranged as follows:

Clause 2 provides definitions of terms and abbreviations.

Clause 3 reviews the traditional approaches to image search and from various examples, motivates the importance of the user in the search process, the importance of making explicit the user task and user evaluation, and that meaning in images may be as much added from outside as extracted from the image itself. This motivates the next clause, which are the real use cases.

Clause 4 describes the real use cases of the various ways searching does take place. In particular, there could be multiple entry points into the overall search framework. These would include automatic, semi-automatic, and human (user) driven searches. These would include specific use cases and motivating examples of searches for images.

Clause 5 describes the overall search and management process. This can be considered the requirements specifications for a general search and management architecture.

Clause 6 describes the 4-layer architecture for JPSearch and explicitly identifies the components in the architecture, and what their roles and positions are in the architecture. We will describe how the use cases described in Section 3 map to the layers of this architecture.

Clause 7 specifies the overall structure of JPSearch.

2 Terms, Definitions and Abbreviated Terms

For the purposes of this document, the following terms, definitions and abbreviated terms apply.

2.1 Terms and definitions

2.1.1
Annotation
metadata added to an image by way of definition or comment

NOTE It is normally in text and done by a human.

2.1.2
Content based retrieval system
system using non-text features of a content to search for similar contents

NOTE The abbreviated term CBIR is used to represent content-based retrieval system for image.

2.1.3
Communal recommendation systems
Systems which track items viewed by customers, and which groups customers by similar interests then recommends items viewed by one customer to another customer in the same group

2.1.4
Context of a user
circumstances and conditions of the user during a query

2.1.5
Context of a query
context of the user as well as other circumstances and conditions affecting the query

2.1.6
Contextualization
process of placing a user or a process (e.g. a search) into a context

2.1.7
Index
way of organizing data that improves searching the data

EXAMPLE A library catalogue is an index. It is normally sorted alphabetically to make it more efficient to find an entry in the index. Indexing is the process of creating an index.

2.1.8
Metadata
data about data

EXAMPLE An image is a data item. Metadata about the image may include information such as the size of the image, the date it was created, etc.

2.1.9
Ontology
model that represents a domain and is used to reason about the objects in that domain and the relations between them

NOTE This is a form of knowledge representation about the world or some part of it.

2.1.10
Pragmatics
leftover part in a theory of language and communication after the syntax and semantics have been taken out

2.1.11
Query
request for information from a search and retrieval system

2.1.12
Query expansion
technique in information retrieval that adds terms to a query to improve the accuracy of the search or to increase the number of useful documents retrieved

2.1.13
Query-by-example
type of query where an example of the answer desired is used as the input to the search system

2.1.14
Reverse Index
type of index where, given a word, one can look up the reverse index and locate all the documents where the word occurs, and optionally, where in the document and how often the word occurs

2.1.15
Semantics
mapping between elements of a language and the real world

2.1.16
Syntax
set of rules that govern whether a sentence (or other unit of communication) is well formed

2.2 Symbols and abbreviated terms

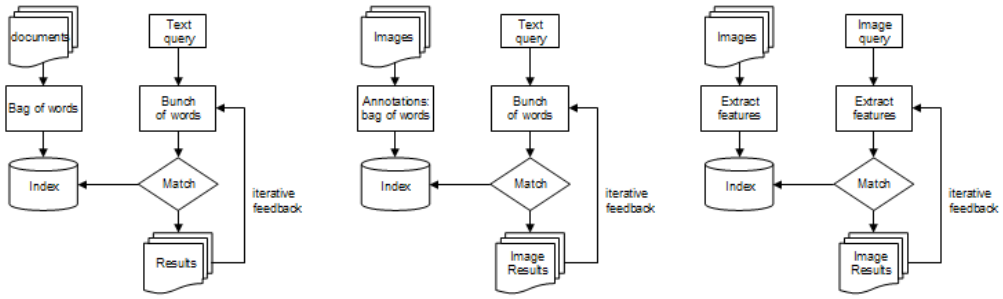
ALT:	Alternative Text
CBIR:	Content-Based Image Retrieval
HTML:	Hypertext Markup Language
QBE:	Query-by-example
UI:	User Interface
USB:	Universal Serial Bus

3 Background and motivation of a user-centric approach to image search

3.1 Traditional Models of Image Search and Retrieval

To put things into perspective, we need to start with the field of document or text search and retrieval. This has a long rich history rooted in the field of library science and has evolved through forms of text processing, information tracking, through to document “retrieval engines”, leading to the various web search engines. Figure 1(a) shows the naive system view for document search and retrieval that made up the early retrieval systems. Documents in the collection to be searched had first to be indexed. This was commonly to treat them as a “bag of words” that were then efficiently stored in a reverse index. A query is normally a bunch of keywords that is matched against the index and the appropriate documents retrieved.

Figures 1(b) and 1(c) show the traditional approaches to digital image searching. This has primarily fallen into two camps. The first, as shown in Figure 1(b), is to search by keywords, i.e., it requires each image to be associated with one or more keywords. There are various schemes to associate the keywords with the images but the most effective has been manual annotation, e.g., done by home users for their home photos or by trained domain experts in more commercial settings. As can be seen from comparing 1(a) and 1(b), it is essentially the same kind of naive document search system.



(a) Traditional Document Search and Retrieval

(b) Image Search and Retrieval using annotations (e.g., keywords)

(c) Image Search and Retrieval Using an image as a query i.e., Content Based IR

Figure 1 — Naive system view of information retrieval, Image search, and CBIR.

The second common form of image retrieval uses an image as a query and the system attempts to retrieve other images that are similar (Figure 1(c)). This is the accepted state of the art in content-based information retrieval (CBIR) systems. The primary area of research is oriented around discovering and extracting new kinds of image features to characterize the image for better performance during retrieval. Some of these features include color (in the form of color histograms, color moments, color deltas, etc.), edges (whether as line features or "assembled" into higher-order objects), texture, blobs, regions, etc. While the features have become more sophisticated, fundamentally the systems follow the model in Figure 1(c). Again as can be seen by comparing with Figures 1(a) and 1(b), there is little difference with the other naive systems.

There are also hybrid systems that combine text and image queries for searching. Text queries are used as an entry point into a search or browse space, after which image-to-image matching is used to refine the query or to retrieve further "similar" results. The system as a whole becomes more complex, but each of the components still behaves as above.

3.2 The user as part of the retrieval system

Figure 1 showed the similarity of current image retrieval models to the naive models of document retrieval prevalent about 20 years ago. Since that time, document and text retrieval has improved by leaps and bounds. One of the clear factors in this evolution is the recognition that the user should be treated as an integral part of the information retrieval process, i.e., that finding the right information is about much more than just the search system provided by a library or by a vendor or on a website.

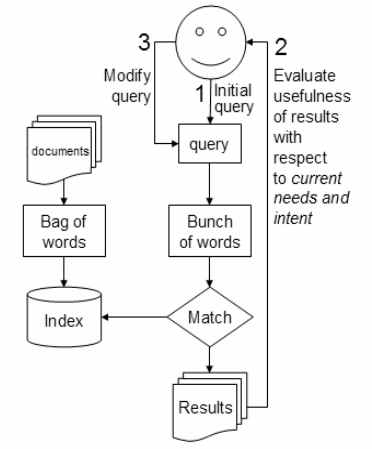


Figure 2 — Adding the user into the system model

Having the user in the loop provides many advantages to a search system. One of the biggest problems in search is ambiguity i.e., which meaning of a word, or which aspect of an image, is the one that is appropriate in a given search situation. Consider a chair. If we were to sit on it, it would be a chair. If we wanted to buy one, we couldn't go to a "chair store"; we have to go to a furniture store. And assuming it's a wooden chair, if we were stuck in a blizzard and feeling really cold, it would be firewood! Different aspects of an object, or a picture, or a document are appropriate in different circumstances. These circumstances could depend on the user's needs (I want to buy one), the context (I'm really cold), the utility (firewood), or any one or more of a myriad other conditions. The converse of ambiguity is redundancy where one thing can be referred to by many names, e.g., the planet Venus was also known as *the Morning Star* and as *the Evening Star*, or your portable computer being a *laptop* or else a *notebook* computer.

Having the user as part of the system gives us the model shown in Figure 2. The indexing side hasn't changed; the query process is a little more complicated. The user has an initial query (step 1) for which the system returns a set of results, just as in the naive system. However, here the evaluation of the usefulness of the results is done by the user (step 2) taking into account his or her own circumstances, etc. If necessary, the query can be modified (step 3) and the search done again.

The user also has other influences on how the search system operates. One important recognition is that there is not just one kind of search; that users engage in different *Information Seeking Strategies* (ISS) [Belkin et.al, 1995]. Four dimensions of ISS were identified, specifically Method of Interaction, Goal of Interaction, Mode of Retrieval, and Resource Considered. To quote,

> "method of interaction, can be understood in terms of the classic distinction between *searching* for a known item and looking around, or *scanning*, for something interesting among a collection of items. The goal of the interaction may be *learning* about some aspect of an item or resource, or *selecting* useful items for retrieval. Furthermore, looking for identified items can be characterized as retrieval by *specification*, while identifying relevant items through stimulated association can be characterized as retrieval by *recognition*. And interaction with *information* items themselves can be contrasted with interaction with *meta-information* resources that describe the structure and contents of information objects". [Belkin et.al, 1995, pg 385]

Each of these dimensions represents a choice by the user as they use the search system. If the system does not support that choice, then the user is, in a sense, fighting with the system to achieve his or her goals. For example, traditional libraries (the ones with books on shelves) easily support scanning (and consequently serendipity) whereas digital libraries (with only e-books) have to come up with alternatives (e.g., communal recommendation systems).

3.3 User task and user evaluation

In the previous section, we have motivated the various reasons for including the user in the search process. This section looks at the reason for a user to go to a search system in the first place. One does not just decide for no reason to go to a search engine and look for something. A user must start with a purpose, something he or she wants done, and in the course of doing it or planning on how to do it, knows (or finds out) that there is missing information that interferes with the successful completion of what is to be done. This has been characterized, for example, as a *state of anomalous knowledge* [Belkin, et. al, 1982]. It is important to note that the user is in a *state*, for this implies a larger context within which one or more pieces of information are missing. The user goes to the search system to find those missing pieces of information. Finding the missing information and thus allowing the purpose to be satisfied is called the *task*.


Web Results 1 - 10 of about 245,000 for eiffel tower restaurant. (0.26 seconds)

Eiffel Tower Restaurant - Las Vegas - Home
French cuisine and picturesque views from the Eiffel Tower Las Vegas. Includes menu.
www.eiffeltowerrestaurant.com/ - 14k - Cached - Similar pages

Paris Pages: Tour Eiffel - Monument
The Eiffel Tower was built for the International Exhibition of Paris of 1889 ...
The tower has three platforms. A restaurant (extremely expensive; ...
www.paris.org/Monuments/Eiffel/ - 25k - 1 Jul 2005 - Cached - Similar pages

Altitude 95 Eiffel Tower - Restaurants - Eiffel Tower-7e - Paris ...
Your guide to Paris, France hotels, restaurants, events, shopping.
www.hiptravelguide.com/paris/reviews-41.html - 45k - Cached - Similar pages

Restaurants
... over the Seine and the Trocadéro to one side and the inside of the Tower to
the other. ... Commercial Service Mail: tour.eiffel.rv@elior.com Restaurant ...
www.tour-eiffel.fr/teiffel/uk/pratique/resto/ - 10k - 1 Jul 2005 - Cached - Similar pages

Paris Las Vegas: Fine Dining
... Chef J. Joho's Eiffel Tower Restaurant serves the most acclaimed French ...
the Eiffel Tower Restaurant provides one of the most vibrant views in all of ...
www.caesars.com/Paris/LasVegas/ Dining/FineDining/EiffelTowerRestaurant.htm - 24k -
Cached - Similar pages

Restaurant Home Page | Restaurants | Lettuce Entertain You
Visit Eiffel Tower Restaurant's Web Site. PHONE NO.: (702) 948-6937. LOCATION:
3655 Las Vegas Blvd. South, Las Vegas NV, 89019 ...
www.leye.com/restaurants/rest_home.jsp?id=30 - 41k - Cached - Similar pages

Figure 3 — Results from a typical search engine for *Eiffel tower restaurant*

A typical search engine returns a list of results. These are often ranked, i.e., the topmost result is what the engine predicts to be the most relevant to the given query. Note that this is with respect to a *query* and not to a user or a task. The change in paradigm that comes with making the user a part of the information retrieval system is that relevance is now defined with respect to the user and the task. This is called *user evaluation*.

Consider the two following scenarios. John is attending a convention in Las Vegas and wants to find the phone number of the Eiffel Tower Restaurant at the Paris Hotel to make a dinner reservation. Betty is a student trying to find the name of the restaurant on the Eiffel Tower (the one in Paris, France) observation deck for a school project. Here we see two different users each with a very distinct task. But when it comes to the search engine, they enter identical query terms. Both unsurprisingly type in "*Eiffel tower restaurant"*. The first 6 results of the search are shown in Figure 3.

If you take the naive system view, i.e., that you evaluate relevance with respect to the query, then all 6 results returned were "relevant" since they were all about *Eiffel tower restaurant* in some way or another. However, when you include the user and the task in the evaluation, this changes very dramatically. With user evaluation, the 1st, 5th and 6th results are relevant to John, and the rest are relevant to Betty, i.e., only 50% of the documents returned are relevant in either case, and none of the documents relevant to John *in his task* was relevant to Betty *in her task*. Incidentally, notice that the 6th and 3rd results respectively already satisfy John's and Betty's tasks even without having to open the documents (websites).