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

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.

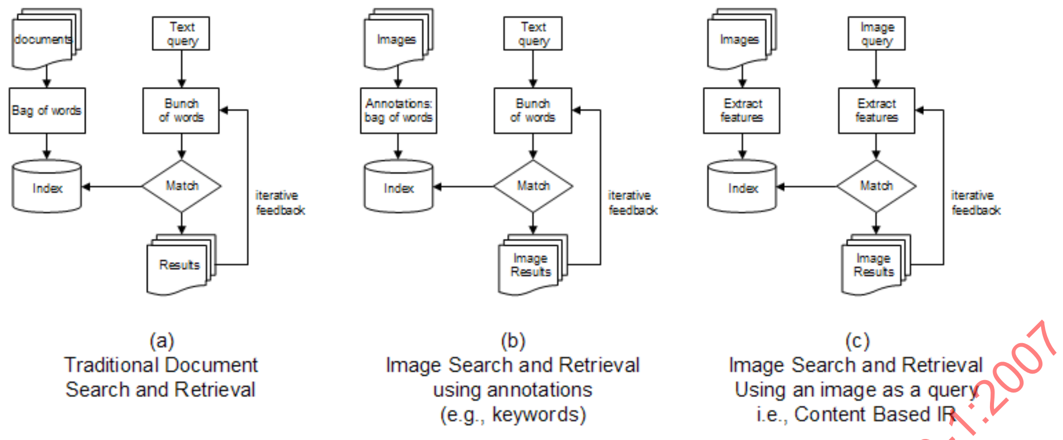

Figure 1 — Naive system view of information retrieval, Image search, and CBIR.

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.

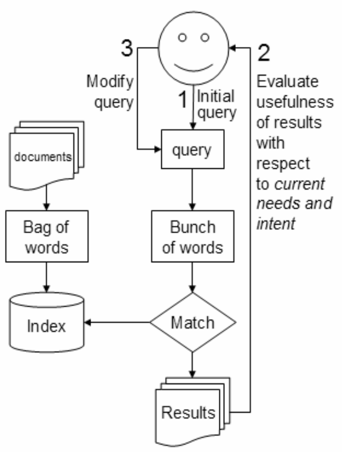

Figure 2 — Adding the user into the system model

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 1^{st}, 5^{th} and 6^{th} 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 6^{th} and 3^{rd} results respectively already satisfy John's and Betty's tasks even without having to open the documents (websites).

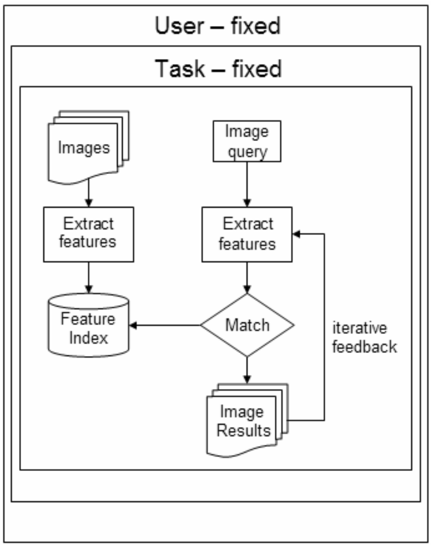

Figure 4 — When you fix the user and the task, the model reduces to that in Figure 1

A text example was used above for clarification, but the idea of user task and user evaluation applies equally well for digital image searching. The state of the art in image searching research does include a user and a task. However, to make things less subjective, researchers hold the user and the task as constants. For a given query, they are then able to decide which images are "relevant" and which are not. For our example above, this would be to say that either the answers for John are "right" xor those for Betty are "right", without considering (and supporting) the possibility that both sets of answers could be "right" in the appropriate circumstances. In image retrieval systems, this fixed user and fixed task evaluation is called the *ground truth*, and is used as an objective standard to measure the performance of an image search engine. While this makes the comparison between different search systems easier, once you fix the user and task, it gives rise to the situation in Figure 4, i.e., the system reduces back to the model in Figure 1 (the naive system view). What this means is that the system measures relatively well for a given user and a given task, but *only* for that given user in that given task. A generic model for digital image searching must be able to handle different users (with their respective problems they want to solve) and different, even conflicting, tasks at the same time.

3.4 Meaning also comes from outside

Semantics is a misnomer in the way it is used in the image (or video) searching community. Linguistically speaking [Newmeyer1986], semantics refers to the mapping between well-formed expressions in a language and the real world. It is invariably compositional, i.e., if I know what *a* and *b* denote and I know what the operator + denotes, then I know what *a+b* denotes. However this doesn't work for image "semantics". Just because I know that this region is a blob, that it's round and that it's red doesn't mean I know that it's a tomato. Or a red ball. Or the sun. Or whatever else could be round and red. In fact, the sun has often been referred to as a red ball, so these are not exclusive possibilities. Properly, the idea of trying to make sense out of an image is closer to what linguists term *pragmatics* rather than *semantics*. Pragmatics is very much what it seems to be [Levinson 1983]. Without worrying about formal definitions of atomic units (pixels? blobs?) or compositional rules (2 adjacent pixels of the same color = a blob of that color?), pragmatics tries to figure out how two people are able to communicate even if they are using (or misusing) a language with nonsense referents and non-well-formed utterances.

Figure 5 — picture taken from www.whitehouse.gov

The distinction between semantics and pragmatics is important. Linguistically, semantics is defined with respect to the utterance only, i.e., it doesn't matter what the speaker of the utterance meant when he made the utterance; nor does it depend on what the listener took away from hearing the utterance. There is, in other words, the same idea of a user-independent ground truth. Pragmatics, on the other hand, is very dependent on the speaker and the listener. Consider Joan and Darby, who are a long married couple and they're at the Cineplex. They're standing looking at the list of movies showing and after a couple of minutes; Darby turns to Joan and says, "So?" That's not a well-formed sentence and has no semantics, but that single utterance, pragmatically, carried a wealth of meaning. Darby, by that one-word utterance, was saying to Joan, "there's your favorite romantic weepy showing that always puts me to sleep, and the usual blockbuster action movie that gives you a headache, so are we going to settle for our normal campy comedy or try out that animated children's film?". This is not an unbelievable scenario by any means. What it shows is that once you have the human user in the loop, it's pragmatics and not semantics that is important, and that pragmatics *needs* the intelligence and the interpretation of the human to make it work. Where there isn't many years of marriage to provide a shared context, there is often redundancy, social norms, and shared objectives that fulfil the same role.

So, this brings us to image information retrieval. In text retrieval, a document is effectively indexed as a bag of words, without reference to outside pragmatics. There are heuristics that are useful, e.g., that the more times a word occur in a document, the more likely it is that the word is a better proxy for the document during a search. In text-based image retrieval (i.e., using annotations), the annotated text is the proxy, and once the annotations are entered, there is again no reference to outside pragmatics. Where the images occur in context, e.g., on a webpage, then the caption, the HTML ALT text for the image, or the text on the page can be and are used as proxies. However, there may be smart schemes to create "better" annotations, or to automatically create annotations for the text. These then require information to be *added* to the image. Consider Figure 5. If the annotation was "man using a telephone", it might be argued that this is a very smart *extraction* of semantics from the image. An even smarter one would be "George Bush in the Oval Office talking on the phone". The actual caption (annotation) of the picture says "President George W. Bush speaks via phone to Associate Supreme Court Justice Sandra Day O'Connor Friday, July 1, 2005, shortly after she submitted her letter of resignation citing personal reasons. The letter sits on the desk." That annotation contains information that obviously cannot be extracted from the image, including (a) that the date is 1 July, (b) that the person on the other end of the phone is Sandra Day O'Connor, and (c) that she had just resigned, etc.

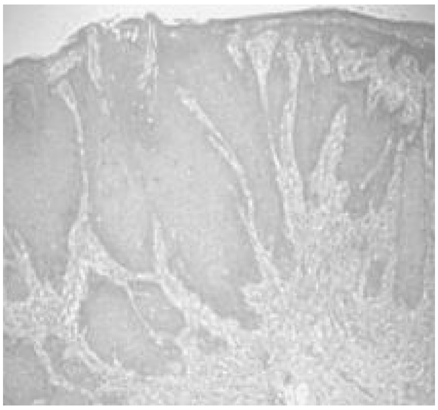

Figure 6 — Microscope slide of oral tissue showing cancer growth

Lastly in content-based image retrieval (CBIR), we have multiple layers of meaning all potentially active at the same time. There are primary features that can be extracted from an image. These are unambiguous information that is structurally (or syntactically) objective such as color histogram, texture, lines, etc. This has also been called low-level features in the CBIR literature. There are also secondary features that can be extracted from an image, especially in specialized domains such as medical image analysis [Muller et.al 2004]. These are features which form is intrinsic to the image, but which significance comes from outside. For example, a particular texture from a tissue image on a medical slide may have a high correlation with cancer (see Figure 6).

Then there are higher level of “semantics” that come from analysis and deduction (e.g., face recognition in images). Looking at Figure 5 again, if a face recognition tool identified the person in the image as “George Bush”, it would require *adding* information from outside the image to the meaning of the image itself. Consider the following text taken from a news source:

> "The president of the United States said in South Korea that the United States has no intention to attack North Korea. They’ve been told they can have multilateral security assurances if they will make the important decision to give up their nuclear weapons program"

If after reading the above statement, I were to summarize the text by saying “George Bush says that the US will not attack N. Korea”, I would also have had to *add* information in that summary, specifically that *George Bush* denotes the same person (at this place and time) as the definite description, *the president of the United States*.

Another kind of higher level “semantics” comes from looking at similarities that occur across images, i.e., within a class of images [Lim & Jin, 2004a, 2004b]. So pictures of beach scenes have similar features because they have similar content. One blue blob in an image may not have any useful meaning. But when similar blue blobs occur in many pictures, then it becomes interesting. We can sometimes give labels to these types of natural classes, e.g., for the class of beach scenes, we have sky, sea, cloud, sand, etc., though there are sometimes patterns in such image collections that the system may discover but which may not correspond to human labels.

In summary, we have shown that there are two ways of ascribing meaning to an image. The simple one is to consider only the intrinsic image content. The more sophisticated version uses a lot of external knowledge, context, or reasoning. In particular, these sophisticated methods may require considering the image content in comparison with other images in a class, or with respect to extrinsic prior knowledge. In these and other cases, meaning is *added* to the image rather than being *extracted* from an image.

4 Use Cases

4.1 Introduction

From the previous clause, we see that the JPSearch framework architecture needs to be able to handle very diverse ways of ascribing semantics (i.e., meaning in the form of metadata, context, profiles, etc.) to images, and also to collections of images. Thus many use cases were solicited from consumer, professional and academic sources. These were discussed and analyzed and the following use cases were deemed to be representative of the wide spectrum of user needs as discussed in clause 4.

The following descriptions of use cases are only summaries. The full description of each use case is provided in Annex A.

4.2 Searching images in stock photo collections for usage in magazines

The user wishes to buy a selection of images in order to illustrate a publication to be sold to consumers.

4.3 Searching for and publishing authoritative themed sub-collections of images

The user has specific directives and procedures to annotate and manage themed collections of digital images, e.g., “Fake Boticelli Paintings”. The collections will then be accessed by third parties (e.g., professional users purchasing material for commercial use). In particular, this refers to creating sub-collections of themes that are not explicitly annotated in the repository.

4.4 Mobile Tourist Information

A user (tourist) is in an unfamiliar place, sees an interesting landmark and wants to know what it is. He takes a picture of the landmark on his mobile phone and sends it to a tourist information server that calls him back and gives him the information.

4.5 Surveillance Search from Desktop to Mobile Device with Alerts

The user sets up a visual surveillance query on a desktop computer (large screen, comfortable keyboard), saves the query for real time monitoring with results saved periodically for retrieval from a mobile device. In addition, a trigger event occurs, a short message alert is sent to the user.

4.6 Ad hoc search without time-consuming housekeeping tasks

Users wish to bring their private photo collection on their personal storage devices such as a memory card and to retrieve images using any terminal, i.e., computers in friends’ homes, by connecting the devices to the terminal.

4.7 Rights clearance to publish a compliant business document

Business documents are often circulated in a company and sometimes their usage conditions are missing. This results in a critical business problem especially if it contains intellectual property of a 3rd party. Users should be able to check that the figures in their business documents are compliant to the usage conditions.

4.8 Tracking an object creation process using a temporal series of photos

In several countries/regions, archiving of the huge amount of photos capturing the construction of a building is mandatory (by law or contract). Thus the user has a temporal series of photos which capture a certain object and the user can check what happened to the object and when it happened. But it is very hard to retrieve these photos in an efficient way.

4.9 Finding illegal or unauthorized use of images

The user holds his original content. He wants to find unauthorized variations of his original content using search engines.

4.10 Finding the best shots or filtering out of the worse shots

The user has a huge photo collection, which includes all his photos taken in a year. At the end of the year, he/she would like to create a greeting card with the best shot.

4.11 Context searching without human annotation

The context is specified (either implicitly and/or explicitly) by the user (e.g. portrait mode (implicit) and Birthday party (explicit)) for a set of captured images. Later on they wish to search their collection using metadata that is generated based on that context (e.g. face detection and recognition).

4.12 Image search based on image quality

A consumer wishes to create a shareable artifact such as a photo calendar using images from his collection that are of high objective image quality and/or that are also of high subjective quality. The system suggests possible candidates to the user.

4.13 Image search with deduplication

A user does a search and retrieves many images with different filenames that turn out to be the same picture, that may be of different formats, sizes, and quality. The user wants to get a unique picture, preferably the legal one with the best quality.

4.14 Matching images between collections for synchronization

Consumers and other users of digital images end up with a large collection of images stored on multiple computers, laptops, external drives, portable USB keydrives, portable photo players, photo sharing websites, photo printing websites, mobile phones, etc. They upload new images to any subset of these. They organize parts of their collection independently. They annotate on some platforms.

Consumers and other users of digital images thus end up with a large collection of images stored on multiple platforms. They need their collection to be synchronized across all platforms including file synchronization, collection structure and metadata synchronization and image metadata synchronization. Given an image on one platform they want to identify the matching image on another platform in order that synchronization can be performed. Also, images are edited and enhanced on different platforms; these different versions need to be reconciled so the user does not repeat work, or create inconsistencies.

This use case is also extended to handle the situation where a user may desire to transfer existing data and metadata to a newer and/or better system; where such a system may be a centralized repository; and where the user may desire to archive data in a format intended to survive current products.

4.15 Social metadata updating and sharing of images for searching

A user, Kim, takes many pictures with people in them. He uses an automatic face recognizer to annotate the pictures with the names of the people in it. However, there are faces that his system does not know even though they occur many times (i.e., not just incidental “noise”). This lack of information may be because Kim does not know the face, or has not entered the information. The image collection and metadata is shared with his family, colleagues and friends (which form his social circle(s)). Many of them also use face recognizers and annotation programs. By pooling their knowledge of faces and names, faces can be recognized and annotated even if Kim does not have the information in his system.

4.16 Image search in the medical domain

Many pathologies have visual symptoms that are essential for doctors doing diagnosis. There are cases where the visual symptoms are not familiar and the doctor would like to consult a visual reference material (an atlas, or previously diagnosed reference cases). The user (a doctor in a hospital clinic) searches them for the best matches to the symptoms and retrieves case histories (including other images, metadata, text, etc.) to aid his/her diagnosis.

4.17 Servants image searchers

The user accesses content by means of a standard web browser. The web site forwards the user’s requests to the application using a standard communication protocol. For each provided query, the application creates a personalized intelligent search agent that will effectively take care of the search on behalf of the user that uses the manager layer to find results in the content provider's archives. The use case is related to professional users who have thematic permanent search activities: news writers, researchers, book writers, etc. The user has his personal profile and can personalize the results into collections and annotate them. Part of the personalized annotations can be shared with staff or community users.

4.18 Open federated repositories

Digital archives are going to set up access to part of their databases for consultation from cultural and institutional search interfaces.

The data provider manages the database or collections locally and provides a standardized protocol to consult the data. The service provider will manage the added value services (search and retrieval and presentation). The service provider also harvests the collection and will provide the user interface respecting the middleware specifications.

The contents that may also self-announce: JPSearch content could be compliant to an existing solution allowing retrieval by a low cost process and performance improvement.

5 Image search and management process

5.1 Introduction

There are two types of data flows in a digital image management system. One is the metadata flow through its lifecycle, that is, creation, maintenance and disposal. And the other is the "query and result" data flow between the user and the database.

5.2 Metadata flow

All the search and management systems use metadata to understand image content. Figure 7 shows the data flow of the metadata.

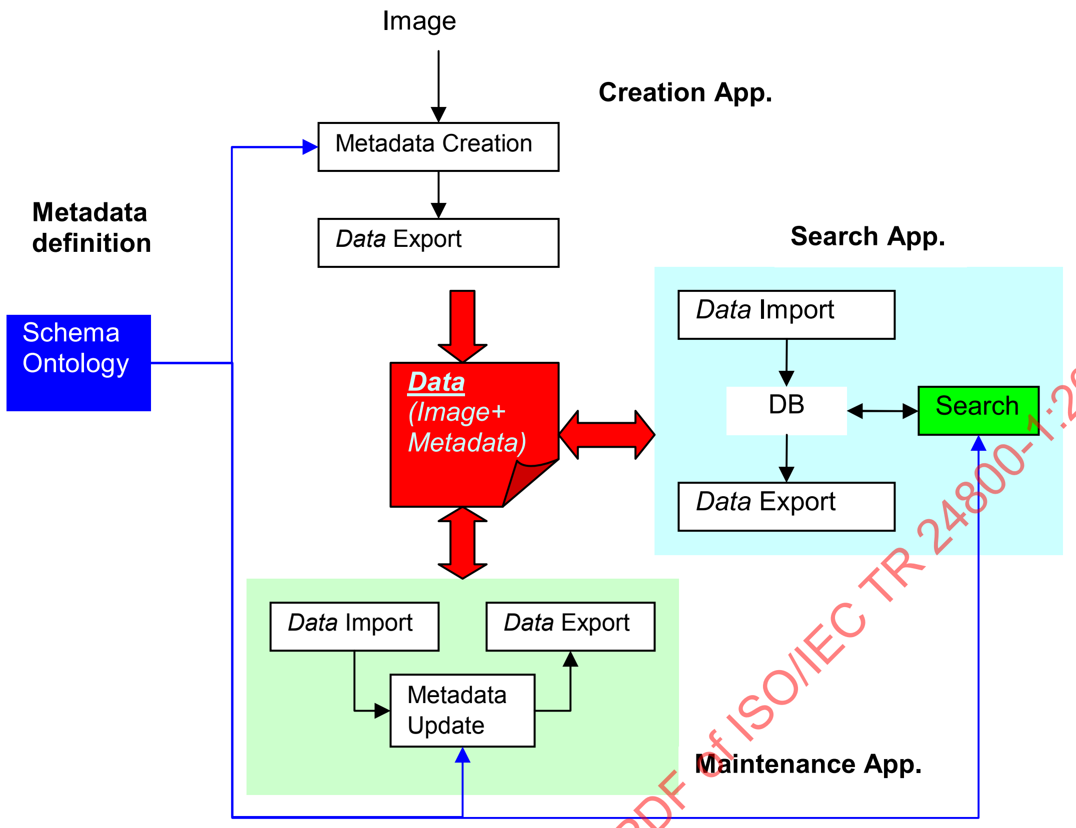

Figure 7 — Metadata flow among image search and management applications.

Three types of application groups that handle metadata exist, that is, indexing, search and maintenance applications. The image data is input to the metadata creation application to create its metadata, which is packed together with the image data and distributed to other systems. The maintenance applications aim to verify the latest status of the image data and their associated metadata and update/modify them if necessary. The detail of the search application is described in the following subclause.

There are two points to be considered to build interoperable search and management systems. One is the semantics definition of the metadata components. Metadata is a set of semantics entities describing what the content is. It can be considered as a projection of pixel sets into a semantics space. Therefore if the semantics of each component is not defined clearly enough, it is hard to understand what the component means. Sharing schema and ontology helps machines to understand the semantics correctly. The Registration of these information and identification scheme shall be standardized. The other important thing is a carriage mechanism to guarantee the persistent association of metadata with content. Since metadata is associated information to the content, it shall be tightly connected to the content and be appropriately updated when the content is modified. This implies that the standardized data exchange format for metadata embedded in image data is needed.

Let us discuss more in detail about each process;

Metadata definition: Schema and ontology selection

(a) we need to decide which features we want to handle in the schema and all the ramifications that arise from that, including the following:

- Decide on features;
- Select corresponding index structures;

- Preferably decide on intercommunication between index structures, e.g., how and where does the evidence from one index structure combine with that from another in the same database.

It is also possible to envision a search framework that is designed so that new features can be added in a “plug and play” manner, and so that the query process can automatically make use of the new capability.

(b) Select one or more ontological schemas to provide structure for indexing and later querying. This would be equivalent, for example, to decide on using the Library of Congress classifications when you are setting up a conventional library. An ontology schema may include inheritance.

Indexing App.: Metadata creation to enable search and management applications

Input, analyze, and process the images. This roughly comprises the following activities for each image, though which steps are done depend very much on a particular system:

- process the intrinsic information: (i) the raw image data, compressed or uncompressed, (ii) the data in the header (e.g., EXIF data in jpeg files), and the housekeeping information (e.g., filename, directory in which it is located, timestamp, etc.)
- extract primary visual features such as color histograms, texture, line segments, etc. from the image data. Primary visual features are distinguished by being unambiguous.
- extract secondary visual features such a regions, faces, etc. Secondary features may vary greatly depending on system training, threshold values, fudge factors, etc.
- deduce additional features or information, i.e., add real world knowledge to primary and secondary features to generate new information
- if the image occurs in a context (such as a webpage), then extract the non-visual features. For web image indexing, for example, these would include text annotations, URL information, etc. These are the information most often used as proxies in non-CBIR searches.

Maintenance App.: Metadata maintenance for its lifecycle management

- Create additional metadata and/or update existing metadata if the image data is modified after metadata creation.

Search App.: Metadata registration to image database

(a) Generate the index (or indices as appropriate)

(b) Handle the aggregate data features if necessary. These features are most probably created not extracted. Some examples include the number of images in the collection, categories, clusters, semantic labels about regions over all or a subset of the images, etc.

(c) Augment data features. This includes manual (human or otherwise) augmentation of the feature information. It may also include user feedback to the indexing structures, or labels, or normalization of data features. Also, metadata about the metadata, e.g., confidence level, provenance, authority, privacy, access control, partitions, hierarchical structures, data integrity, etc.

5.3 Query process flow

Figure 8 depicts the search query workflow, within its frame, we considered also the user, the result and the content parts although these three are not the focus of this section.

The relevant layers of the query workflow in a search process are:

➢ Interface (UI) Framework interface

- Query transform
- Execution

These three differentiate each other at the abstraction level, from user interfacing and personalization to the parsing and transformation of the query in a machine intelligible format and then, to the elaboration of the query for retrieving the results and managing the convergences of eventual loops (that could have taken place due to relevance feedback and user interaction).

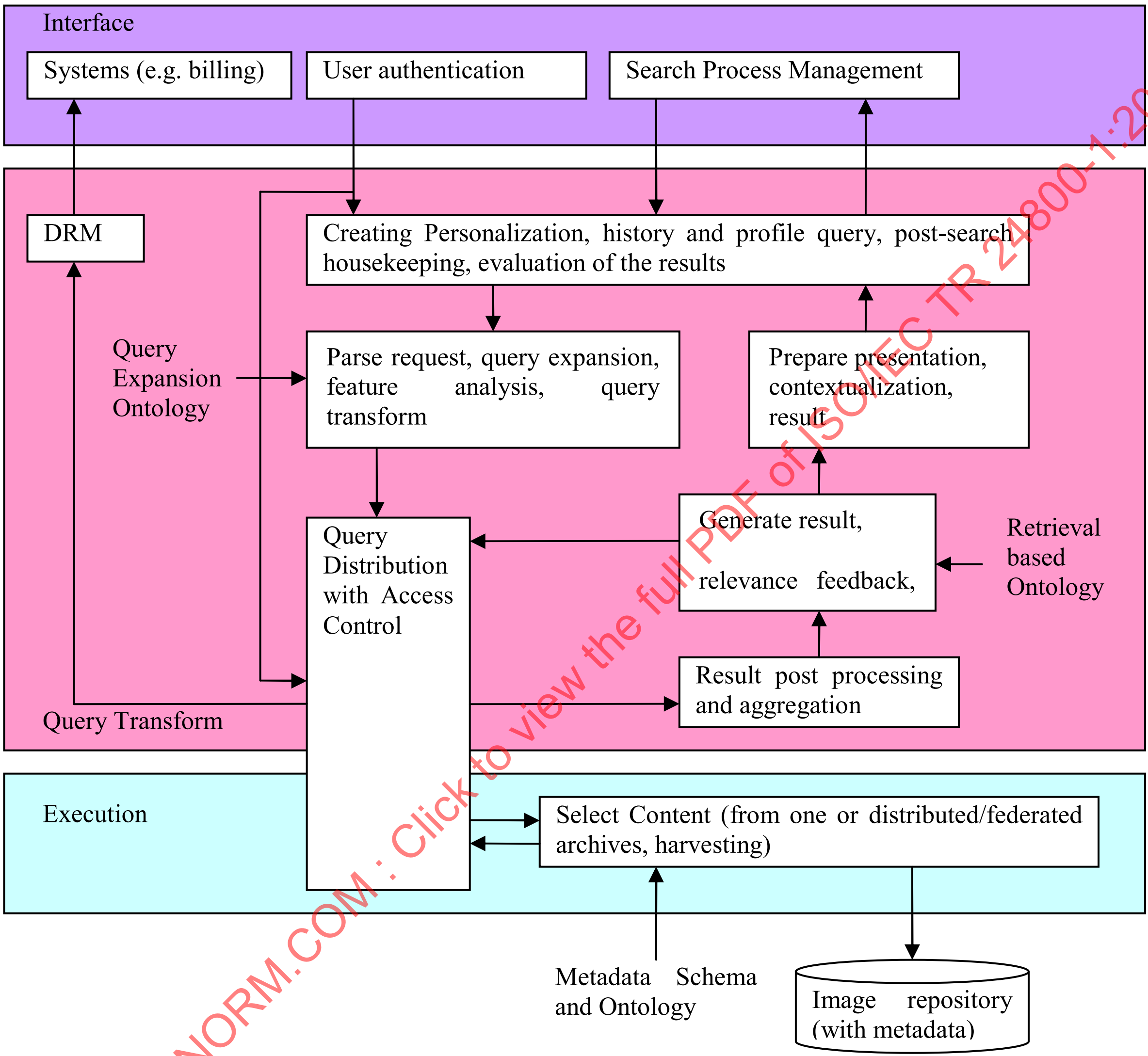

Figure 8 — Query data flow in a search system.

For the query process flow, we assume the presence of a user and the role of the user in the task and in the evaluation of results as motivated above. As a reminder, the user starts with a purpose (a need for information) and wants to accomplish this purpose through the use of an image search system able to converge to his/her needs. This gives rise to the following steps:

(a) The user sets the task. We have described the concept of the task above, and in most cases, there is no explicit definition or articulation of the task. However, there are at least two ways that the task has been built into existing search systems (though text ones and not image ones):

- The user's search history in the current search session. This assumes that the user is still engaged in the same task during the same search session. Systems are able to make use of queries and results from previous searches to augment upcoming queries or to filter results, etc.

- As mentioned above, the task is a reflection of the user state (i.e., context) and there are systems that attempt to elicit the user context to optimize the search process. This may include location (e.g., looking for traffic information, or in the case of images, for maps), time, etc.

(b) Translate the task to the appropriate query for that system. This is circumscribed by what types of querying may be supported by a given system. If multiple query types are supported, then the task (implicitly or otherwise) often helps to decide which type to use. Some examples include text (e.g., abstract concepts in stock photos), or Region-of-Interests (e.g., face search), or highly compressed images (e.g., query on a mobile handset).

(c) Enter the query into the system. This is non trivial. For example, I may have a very clear picture in my mind, but I can't draw very well, so if I have to draw my query then what ends up as input into the system bears little resemblance to the query in my mind. The query modality may include the traditional keyboard & mouse, pen interface, microphone, a webcam, etc., and could take many forms including:

- query by example (also called QBE): this uses one or more image as a query.
- query by sketch: similar to QBE, but the user draws the initial input
- query by text: using text as a proxy for the image, e.g., most current internet image search engines
- query by emotion, aesthetics, etc.
- fusion of one or more of the above, optionally with user modification
- each of the methods above could be done as positive input ("like this"), negative input ("not like this"), partial or whole, compositional, etc.
- query by context: the user pastes a long sentence and the system provides adherent results and selectable directions for converging the search.

(d) Pre-search processing. This may involve procedures and information not exposed to the user by the search system, and may include query parsing, representation, translation, transmission, refinement, expansion, normalization, aggregation, parsing the user history and profile, etc.

(e) Execute search. There are many system dependent implementations for search, but common issues include efficient query methodologies, optimization for high precision vs. high recall, tradeoffs between speed, cost, memory, accuracy, relevance, ease-of-use, etc.

(f) Post-search processing. These include:

- Automatic (pseudo) relevance feedback and other search enhancement techniques, then search again. These are automatic techniques that do not require input from the user.
- Handling result metadata, e.g., which databases returned results, in what format, how many, date of the document (image), cost of retrieving document, etc.
- Ranking in one or more dimensions, by relevance or other feature (e.g., cluster by genre, colour), etc.
- Fusing multiple result lists, or evidence from multiple lists, or from intermediate/partial results
- Tradeoffs due to cost, speed, physical limitations of the query device, network, security, privacy, etc.

(g) Retrieval of result lists, and output to the user, e.g., as thumbnails, region-of-interest thumbnails, non-image proxies, other forms of visualization, etc. Consider, for example, if the user was doing the search on a mobile phone vs. doing the search on a desktop computer; the way the result list is presented would be very different.

(h) Evaluation of the results by the user. This was discussed above, but also includes factors such as navigation (how to view the images), operational issues (e.g., authorize payment for paid search results) and two optional activities:

- query refinement by the user based on the current list of results, and then search again. Query refinement includes query expansion, enhancement, modification, etc.

- relevance feedback based on the current list of results and then search again. Relevance feedback is where the user can rate the usefulness (relevance) of a particular result (image) to his or her task. The rating is used to modify the query before running the search again.

The evaluation of the results influences the convergence of the retrieval process: the user can be helped by the system to reduce the result to the most relevant results but he/she leads this process.

(i) Retrieval of result images. This is again non trivial because a user has a particular purpose in searching for the image or images, and this purpose could put constraints on the retrieval of the result images. For example, the user may want a low resolution web-quality image whereas what is available is a high resolution print quality stock photo, or the user may want an authenticated image as a guaranteed reference.

(j) Post-search housekeeping. Many systems offer the option for some kind of store and re-run facility, or an automatic scheduled monitoring service, etc. These require user registration and login, possibly context and a profile, as well as query and result management and storage.

6 JPSearch Architecture

6.1 Overview of the architecture

Figure 9 shows overall architecture of JPSearch. JPSearch requires a 4-layer architecture including a user layer, a query layer, a management layer and a content layer. It consists of 5 independent components:

(a) Query Process
It is located in the query layer in the architecture. It provides users an “intelligent agent” service executing search tasks efficiently. It receives search tasks from the users, expands them, and forms a machine understandable set of queries that is delivered to the lower layer. It also behaves as an intelligent filter for the received results. It validates the results and ranks them according to the users’ evaluation criteria. It may also create a new query (based on relevance feedback) if it is necessary or is requested by the users to improve the results.

(b) Repository Management Process
It is located in the management layer in the architecture. It enables users to search distributed image repositories using the same query. It will deliver the query to the appropriate image repositories and aggregate the results from them.

(c) Image Repository
It is located in the content layer in the architecture. It provides basic search functionality, including receiving a set of queries, performing comparisons between the queries and the stored content metadata, and creating a result.

(d) Metadata creation process
It is located in the content layer in the architecture. Search processes assume that content metadata, i.e., a machine understandable representation of the content semantics, is available. This process provides the functionality to build metadata using a proper schema and ontology definition.

(e) Metadata maintenance process
It is similar to metadata creation process, but provides the functionality to add/replace/remove all/partial metadata entities. Since metadata is associated information to the content, this process needs to verify if the metadata is appropriately updated after the modification of the content (image data).

Note that the user layer aims to personalize the search service. It is considered that its implementation is out of JPSearch's scope.

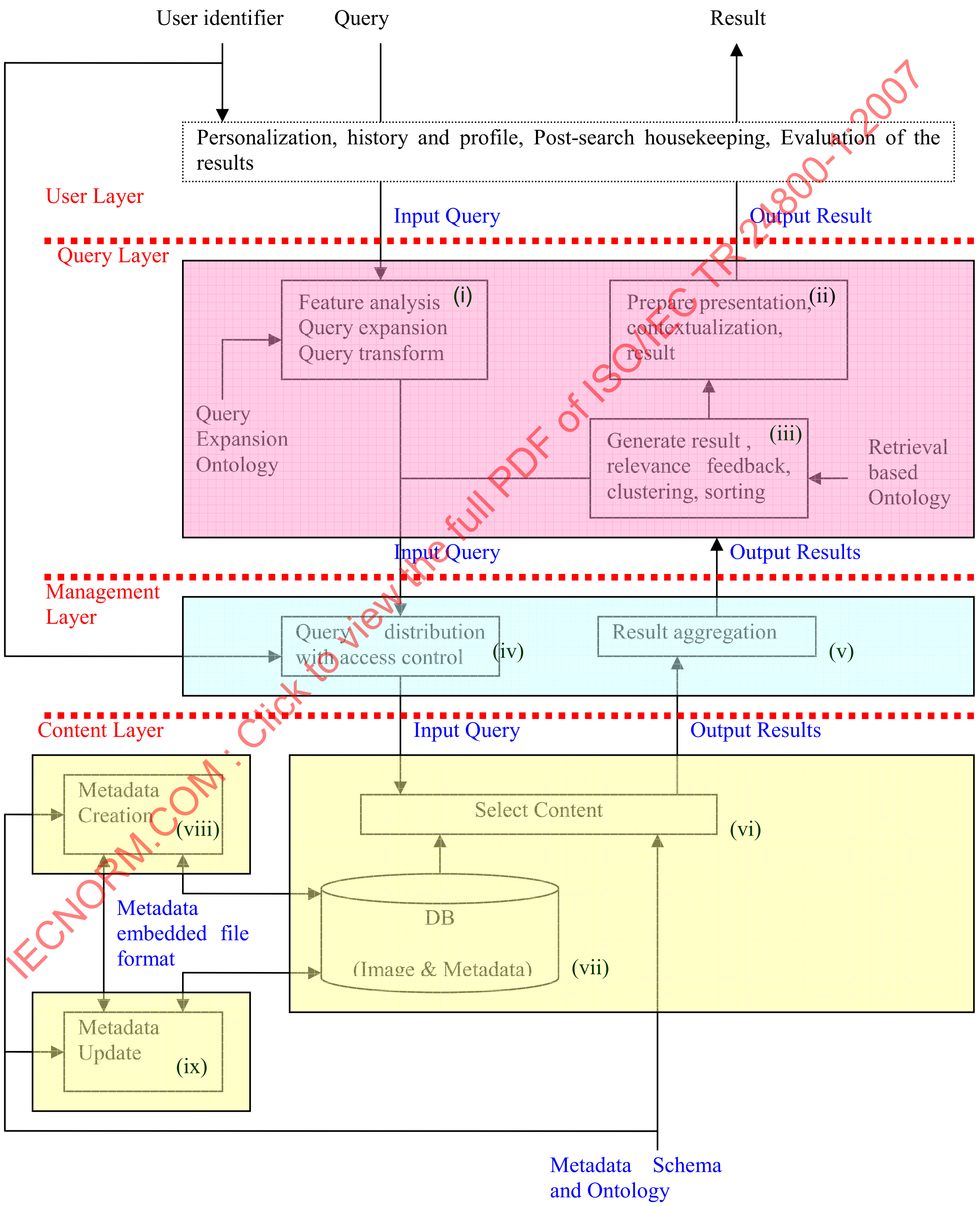

Figure 9 — JPSearch Architecture

6.2 Parts to be standardized

JPSearch specifies the data exchange mechanism between the JPSearch components. The data movement between components in the JPSearch framework is determined based on the process flows described in Clause 5. These are summarized in Table 1 where the data movement is explicitly captured as “Data From” and “Data To”. These are clustered into the five logical groups that represent the items to be standardized in JPSearch. These items are lastly mapped to the upcoming parts of JPSearch, that are described in Clause 7.

Table 1 — Items to be standardized.

JPSearch Parts	Items to be standardized	Data From	Data To
Part 2	Metadata schema and ontology sharing	Repository of schemas and ontologies	Image repository Metadata creation process Metadata maintenance process
Part 3	Input query format	User	Query process Repository management process Image repository
		Query process	Repository management process Image repository
		Image repository	Image repository
	Output result format	Image repository	Repository management process Query process User
		Repository management process	Query process User
		Query process	User
Part 4	Data format (Image+Metadata)	Metadata creation process	Image repository Metadata maintenance process
		Metadata maintenance process	Image repository Metadata maintenance process
		Image repository	Metadata maintenance process Image repository
Part 5	Metadata format (for repository synchronization, upgrading, archiving, etc)	Image repository	Image repository

6.3 Architecture justification with use cases

All the use cases described in Clause 4 can be implemented using JPSearch architecture. Table 2 illustrates what modules are necessary to implement each use case. Here the module numbers (columns) refers to the ones in Figure 9 and the use case numbers (rows) refers to the use cases described in Annex A. An “x” in Table 2 indicates that that module is mandatory to implement the use case. An “(x)” indicates that the module is mandatory in some instances of the use case.

Table 2 — Required modules to implement each use case

Use Case #	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)
#1	x	x	x	(x)	(x)	x	x		
#2	x	x	x	x	x	x	x		x
#3	x	x	x	x	x	x	x		
#4	x	x		(x)	(x)	x	x		
#5				x	x	x	x	x	(x)
#6	x	x		x	x	x	x	x	
#7	x	x	x	(x)	(x)	x	x	x	
#8	x	x	x	x	x	x	x		
#9	x	x	x	(x)	(x)	x	x		
#10	x	x	x	(x)	(x)	x	x	x	x
#11	x	x	x	(x)	(x)	x	x		
#12	x	x	x	x	x	x	x		
#13						x	x		
#14							x	x	x
#15	x	x	x	x	x	x	x		
#16	x	x	x	x	x	x	x		
#17				x	x	x	x		

7 Organization of the JPSearch specification

7.1 Overall Structure of JPSearch

Figure 10 shows the overall structure of the JPSearch standard. It is based on the summary in Table 1. JPSearch is designed as a multi-part specification. Part 1, i.e., this document is a type-3 technical report and provides a global view of the JPSearch specification. In the following parts of ISO/IEC 24800, several components will be specified.

7.2 Part 2: Schema and ontology registration and identification

The scope of the Part 2 is as follows:

- Propose a registry of ontologies that can be imported into a JPSearch compliant system, and standardize a hardware and software independent format to register and manage ontologies within the register.
- Standardize a hardware and software independent format for the import, export and exchange of ontologies, sub-ontologies, and ontology elements. The provision of a standard format for handling the metadata facilitates interoperability between different modules in a JPSearch compliant system, and/or between different JPSearch compliant systems. It also facilitates the independent provision of ontologies to be used by such modules and systems.
- Standardize basic functions to query and manipulate one or more ontologies in a repository of registered ontologies. This repository is not within the scope of JPSearch. The functions will be explored and clarified in this part of the standard, and may include the following functions:
    - Query an ontology to retrieve elements or sub-sets from an ontology or from other relevant ontologies
    - Searching for elements within ontologies and across ontologies
    - Searching for relevant ontologies within an ontology registry
    - Selecting a subset of an ontology

JPSearch Part 2 should fulfill the requirements listed in Part 1 in order to support the use cases. The scope of Part 2 should include the exploration and reference to other standards defining, managing and using ontologies.

7.3 Part 3: JPSearch query format

The scope of work of the JPSearch Query Format (JPQF) framework is to provide three standardized functionalities between users and image repositories. The first functionality is to allow users/systems with a set of precise input parameters to describe their search criteria. The second functionality is to allow users/systems with a set of output parameters to describe the aggregated return result sets for user presentation or machine consumption. The third functionality is the query management process such as relevance feedback. The goal of this work is to define a query language that provides the industry with a standardized format to accept and respond to user/system specification for image searches.

7.4 Part 4: Metadata embedded in image data (JPEG and JPEG 2000) file format

This part specifies image data exchange format with associated metadata to accelerate the re-use of the metadata, which is one of the most important motivation of ISO/IEC 24800 (JPSearch) specifications. It supports two functionalities; mobility of metadata and persistent association of metadata with image. The Part 4 compliant data can be circulated without any information loss among any JPSearch compliant devices and platforms on the content layer, such as metadata creation/update devices and image repositories, so that users are not requested to repeat time-consuming image annotation tasks. It also benefits the developers to minimize their effort to implement metadata updating functions when the image data is modified by existing image editing tools.

This format is a key data to implement metadata flow shown in subclause 5.1 and supports the use cases explained in subclause 4.14 as well as 4.4 and 4.5.

7.5 Part 5: Data interchange format between image repositories

The scope of this part is to standardize a format for the exchange of image collections and respective metadata between JPSearch compliant repositories. The metadata may be at the level of the image or an image collection. By providing a solution for the carriage of image collections and associated metadata between compliant devices and platforms, the JPSearch data interchange format enables the synchronization of repositories in order to facilitate simple and fully interoperable exchange across different devices and platforms. Therefore, it allows leveraging the generally high cost of creating metadata.

The JPSearch data interchange format should fulfill the requirements listed in this document in order to support the use cases.

With the JPSearch data interchange format, data could be easily and faithfully transferred between different hardware and software system. This would facilitate, for example:

- Exchanging data between JPSearch repositories on different devices and platforms
- Consolidating metadata generated on different systems.
- Transferring data to a newer and better system. Consolidating selected data to a centralized repository.
- Archiving data in a format that will survive current products.

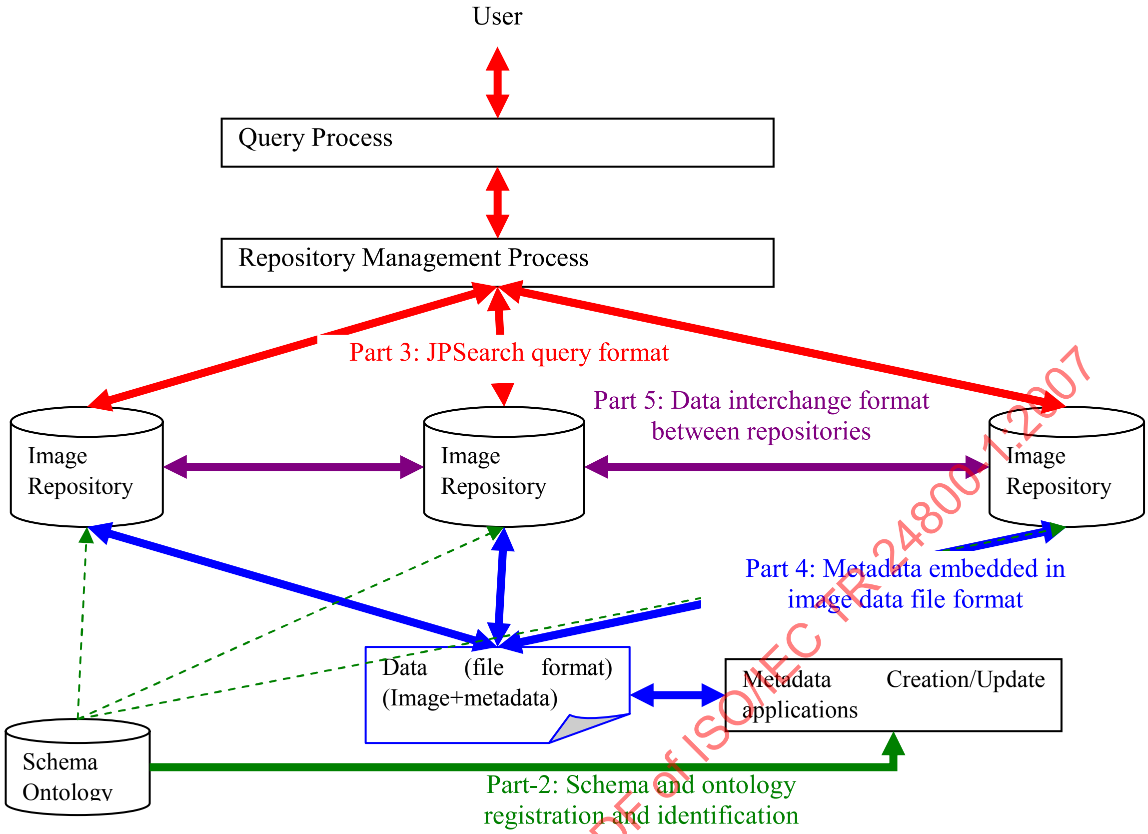

Figure 10 — Overall structure of ISO/IEC 24800 (JPSearch)

Annex A
(informative)

Use Cases

Use Case #1	**Searching images in stock photo collections for usage in magazine**
Summary	The user wishes to buy a selection of images in order to illustrate a publication to be sold to consumers
Description	The user wishes to buy a selection of images in order to illustrate a publication to be sold to consumers.
Application Domains	Magazine editing, Trade journal publication, Corporate website/print publication/marketing collateral branding, etc.
Issues – current problems and benefits	1. Distributed image search – search across multiple image stock photo houses using the same query and combining results – requires a standard. 2. Search based on emotions/mood/theme is hard if there is insufficient textual metadata. It is hard especially if the set of human annotations indicating mood, emotion etc. differ across stock photo houses. 3. Finding a collection of images that best satisfies the query, where the collection is homogenous in some way (e.g. same photographer, same style, same photo house (multi-image discount)), rather than finding single images, requires query specification in terms of collection. 4. Clustering the result-set into themes is hard for visual or emotional themes
User & Context	The magazine or publication editor searching for images given a story text or theme
Input	1. Entire text of the story 2. Keywords + keyimages for query by example, e.g. "images that follow the theme of this cover image found" 3. Image regions 4. Relevance feedback on single items and on groups
Output	Collection of images and associated metadata including collection semantics (properties of the entire collection as a whole)
Relevance to JPSearch	1. federated (or distributed) search 2. emotion/mood content-based search (as opposed to images annotated with emotion text labels) clustering image results by ad-hoc themes

Use Case #2	**Searching for and publishing authoritative themed sub-collections of images**
Summary	The user has specific directives and procedures to annotate and manage themed collections of digital images, e.g., "Fake Boticelli Paintings". The collections will then be accessed by third parties (e.g., professional users purchasing material for commercial use). In particular, this refers to creating sub-collections of themes that are not explicitly annotated in the repository.
Description	The user has specific directives and procedures to annotate and manage themed collections of digital images, e.g., "Fake Boticelli Paintings". The

	collections will be then be accessed by third parties (e.g., professional users purchasing material for commercial use). In particular, this refers to creating sub-collections of themes that are *not* explicitly annotated in the repository.
Application Domains	Publishing, Archiving, Curator-ship (of museum, etc.), Scholarship
Issues – current problems and benefits	1. Search is the first step in creating such a sub-collection; recall (completeness) is most important since the results are likely to be refined, annotated and cleaned up by the user, i.e., trade precision for recall. 2. Combination of search terms in words, that are usually very specific e.g. author/location, as well as images to be used for query by example and/or image regions that could be used for a combined visual-semantic search. 3. The visual search is particularly useful when part of the image repository has not yet been catalogued (there are more images than metadata) or when a monitoring activity is planned (retrieving images that have been annotated but that are not retrieved by textual queries).
User & Context	The collection curator or publisher has a new theme and wants to publish an authoritative database (sub-collection) of images in that theme.
Input	1. Theme description 2. Very specific metadata 3. Images and image regions as exemplars 4. Relevance feedback on single items and on groups
Output	Collection of images and associated metadata
Relevance to JPSearch	1. Very high recall search task 2. Authoritative search results

Use Case #3	**Mobile Tourist Information**
Summary	A user (tourist) is in an unfamiliar place, sees an interesting landmark and wants to know what it is. He takes a picture of the landmark on his mobile phone and sends it to a tourist information server that calls him back and gives him the information.
Description	A user (tourist) is in an unfamiliar place, sees an interesting landmark and wants to know what it is. He takes a picture of the landmark on his mobile phone and sends it to a tourist information server that calls him back and gives him the information.
Application Domains	Tourism, Education
Issues – current problems and benefits	1. This is an image retrieval scenario where an image is used to search for an object (the tourist landmark). 2. The backend system needs to be able to match the query image against possible images (one or more) in its database. The problem is that there is no normalization in the query image, i.e., the user can be standing anywhere looking at the landmark and the angle and distance from the landmark may be very different from the reference image(s) in the database. The time of day (i.e., lighting, day/night, etc.) is also an issue. 3. The object to be identified (signal) also has to be separated from its background (noise). 4. This search needs high precision. The system has to get the correct match at the top of the result list otherwise the wrong information will be sent to the user. Recall is irrelevant. 5. The system cannot just return a ranked list. It is better to return no answer

	(i.e., "don't know") then to give the wrong answer. 6. The system must return a result within a given time otherwise the context changes and the result is no longer relevant.
User & Context	Tourist or student wants to find out information about landmarks. May be extended to non-landmarks (e.g., botanical information in a park, etc.)
Input	1. image
Output	Either "don't know" or identify the object in the image
Relevance to JPSearch	1. image to text (identity) search 2. very high precision with no recall requirements 3. time and context constraints (once the user walks on, the query may no longer be meaningful) 4. difficult context based issues (normalization, salient object, foreground/background separation, lighting and environment conditions)
Use Case #4	**Surveillance Search from Desktop to Mobile Device with Alerts**
Summary	The user sets up a visual surveillance query on a desktop computer (large screen, comfortable keyboard), saves the query for real time monitoring with results saved periodically for retrieval from a mobile device. In addition, a trigger event occurs, a short message alert is sent to the user.
Description	The user sets up a visual surveillance query on a desktop computer (large screen, comfortable keyboard), saves the query for real time monitoring with results saved periodically for retrieval from a mobile device. In addition, a trigger event occurs, a short message alert is sent to the user.
Application Domains	Surveillance
Issues – current problems and benefits	1. A typical filtering task comprises a one time query that is run many times. This is a one time visual query that is run continuously. 2. How to specify a grammar for visual events. This is currently done only in research labs. 3. Summary of search results (events detected) sent to the mobile device. 4. Threshold events generate alerts
User & Context	Surveillance administrator setting up an automated surveillance monitoring system.
Input	1. Visual search grammar. This uses an event grammar. 2. Relevant images capturing target events
Output	Summary of events, or alerts
Relevance to JPSearch	1. specifying visual events as a query term 2. federated search over multiple surveillance sources 3. alerts
Use Case #5	**Ad hoc search without time-consuming housekeeping tasks**
Summary	Users wish to bring their private photo collection on their personal storage devices such as a memory card and to retrieve images using any terminal, i.e., computers in friends' homes, by connecting the devices to the terminal.

Description	Users wish to bring their private photo collection on their personal storage devices such as a memory card and to retrieve images using any terminal, i.e., computers in friends' homes, by connecting the devices to the terminal
Application Domains	Personal or enterprise photo library
Issues – current problems and benefits	1. There is no standardized way to exchange the metadata embedded in the picture data. This results in a lack of metadata portability although its creation cost is extremely high in general. 2. A user can plug-in his memory card to a library system on any computer. The library system provides immediately image retrieval function without performing indexing processes involving feature extraction and text annotation. 3. The system is also expected to offer automatic content categorization mechanism to assist photo browsing. 4. These search-related functionalities are offered not only among the private photos but also among the integrated ad-hoc database.
User & Context	The consumer who wants to bring his private photo data with him.
Input	Not Applicable. A conformant indexing system shall provide the file (or a standardized exchange format compliant code-stream) creation function from pure resource (image or collection of images) and its associated metadata.
Output	Not Applicable.
Relevance to JPSearch	1. a standardized file format/data exchange format to embed metadata within a resource 2. well-defined semantics of metadata entries

Use Case #6	**Rights clearance to publish a compliant business document**
Summary	Business documents are often circulated in a company and sometimes their usage conditions are missing. This results in a critical business problem especially if it contains intellectual property of a 3rd party. Users should be able to check that the figures in their business documents are compliant to the usage conditions.
Description	Business documents are often circulated in a company and sometime their usage conditions are missed. This results critical business problem especially if it contains intellectual property of a 3rd party. Users should be able to check that the figures in their business documents are compliant to the usage conditions.
Application Domains	Creation of business documents
Issues – current problems and benefits	1. Allow tracking of the figure creation process, i.e., a user/machine can understand from where components of the figure come. 2. The usage conditions (rights information) for each component are associated as metadata. They may also be embedded inside the final image or figure.
User & Context	Business document creators
Input	Any image or graphic used in a business document

Output	Rights information about the components used in the figure
Relevance to JPSearch	1. Image search using component-based descriptions 2. Component-based Rights descriptions

Use Case #7	**Tracking an object creation process using a temporal series of photos**
Summary	In several countries/regions, archiving of the huge amount of photos capturing the construction of a building is mandatory (by law or contract). Thus the user has a temporal series of photos which capture a certain object and the user can check what happened to the object and when it happened. But it is very hard to retrieve these photos in an efficient way.
Description	In several countries/regions, archiving of the huge amount of photos capturing the construction of a building is mandatory (by law or contract). Thus the user has a temporal series of photos which capture a certain object and the user can check what happened to the object and when it happened. But it is very hard to retrieve these photos in an efficient way.
Application Domains	Production process validation
Issues – current problems and benefits	1. Arranging/clustering photos on the time axis to assist browsing by users 2. The system automatically identifies the difference between the neighboring photos or neighboring photo clusters.
User & Context	Architects, Product owners
Input	5. A series of photos with evaluation point descriptions 6. An image with ROI 7. A pair of photos or photo clusters
Output	a. A set of temporal clusters. Each of the group consists of similar photos from the view point of identified evaluation criteria. b. An image where the specified changes are found c. Differentiated regions
Relevance to JPSearch	1. ROI-based image search 2. Automatic image clustering

Use Case #8	**Finding illegal or unauthorized use of images**
Summary	The user holds his original content. He wants to find unauthorized variations of his original content using search engines.
Description	The user holds his original content. He wants to find unauthorized variations of the original content using search engines
Application Domains	Rights protection
Issues – current problems and benefits	1. The internet is a favourite platform to distribute content, but has the drawback that it is hard to deal with unauthorized usage of the content. An automatic observation system may be able to detect infringements. 2. The illegal contents might be modified, such as color changes, cropping, re-encoding, in order to hamper discovery by the originator. These malicious attacks are hard to be detected using general image search

	technology.
User & Context	Content providers, amateur photographers
Input	1. image signature 2. image itself (original or one of its variations)
Output	1. A list of images found on the network, whose signatures are very similar to the query. 2. A list of images found on the network, whose origins are considered to be the query image.
Relevance to JPSearch	1. An image signature that is robust to various attacks. 2. JPSearch compliant search systems shall accept the query that is provided as the signature

Use Case #9	**Finding the best shots or filtering out the worse shots**
Summary	The user has a huge photo collection, that includes all his photos taken in a year. At the end of the year, he/she would like to create a greeting card with the best shot.
Description	The user has a huge photo collection that includes all the photos he took during a past year. At the end of the year, he/she would like to create a greeting card with the best shot.
Application Domains	Private photo collection management
Issues – current problems and benefits	1. There are several established procedures for taking a cool photo; however, there are no established mechanisms to evaluate photo(s) according to the well-known practices. 2. The system can identify the best shot on the basis of a description (reasoning or evaluation criteria). It is expected that several photos are listed, each corresponding to certain criteria. 3. The system can filter bad shots to reduce the user's effort in manually finding the best shot (if the number of candidate photos is significantly reduced, the time to browse through the pictures will be much shorter) 4. The latest digital still cameras have a function to take several pictures at a time so as not to miss the best one; however, it is the users' responsibility to select one of them. If the users do not perform the selection, they cause a huge overhead since a shot (that consists of several photos) needs a lot of storage.
User & Context	Consumer users
Input	1. A set of photos (entire collection or part of it (sub-collection)) 2. Evaluation criteria description
Output	1. A set of best shots or a set of photos excluding bad shots. 2. Descriptions about evaluation criteria
Relevance to JPSearch	3. Image sorting on several evaluation criteria (multi dimensional)

Use Case #10	**Context Searching without human annotation**
Summary	The context is specified (either implicitly and/or explicitly) by the user (e.g. portrait mode (implicit) and Birthday party (explicit)) for a set of captured images. Later on they wish to search their collection using metadata that is generated based on that context (e.g. face detection and recognition).