People's Democratic Republic of Algeria

Ministry of Higher Education and scientific Research

University kasdi Merbah- ouargla

Faculty of New Information and Communication Technologies
Department of Electronics and Telecommunications

Thesis Submitted to the Department of Electronics and Telecommunications in Candidacy for the Degree of Master in Telecommunications, Option "Telecommunications Systems".

Presented by:

ABID ANFEL & TOBCHI ABIR & FENTIZ OUMNA

theme

Query Expansion using Semantic Similarity in Image Retrieval

Jury Members

Mr. CHENINA HACHEMI	President	U.K.M Ouargla
Dr. YOUCFA ABDELMADJID	Supervisor	U.K.M Ouargla
Dr. CHLAOUARACHID	Examiner	U.K.M Ouargla

Academic Year:2020-2021

شكر وتقدير

الحمد و الشكر لله أولا وأخرا الذي وفقنا وأعاننا على إنهاء هذا البحث والخروج به بهذه الصورة المتكاملة.

ومن باب انه لا يشكر الله من لا يشكر الناس فإننا نتقدم بجزيل الشكر للأستاذ والدكتور **يوسفة عبد المجيد**الذي رافقنا في مسيرتنا لانجاز هذا البحث وكانت له بصمات واضحة من خلال توجيهاته وانتقاداته البناءة لولاه لم يكن بحثنا ليصل إلى أفضل ما هو عليه الآن . كما نشكر عائلاتنا التي صبرت وتحملت معنا ورفدتنا بالكثير من الدعم على جميع الأصعدة . ونشكر الأصدقاء والأحباب وكل من قدم لنا الدعم المعنوي من قريب أو بعيد.

الإهداء

الحمد لله الذي بنعمته تتم الصالحات.

أولا اهدي فرحة تخرجي وثمرة جهدي وتعبي لنفسي لأني استحق ذلك أستحق النجاح وفخورة جدا لما أنا عليه الآن

السعادة في عين أبي وأمي تعني لي الدنيا كلها الخير فيهم وأكرمني الله بسبب بركتهم التي تحيط بي كل الشكر لكم يا نور حياتي .وأشارك فرحتي مع كل إخوتي وأصدقائي وأقاربي دمتم لي شيئا جميلا

لكم مني كل الحب والاحترام .

عبيد أنفال

الإهداء

نحمد الله و نشكره على أن جعلنا من طلاب العلم و على ما من به علينا من إنتهاء من هذا العمل

و انطلاقا من قولي النبي صلى الله علية و سلم" (لا يشكر الله من لا يشكر الناس")

أهذي ثمرة جهدي المتواضع

إلى قرة العين الى من جعلت الجنة تحث قدميها , الى التي وهبتني كل شيء الى أغلى إنسان أمي الغالية

حفظها الله

الى الرجل العظيم ,صاحب الصبر الجميل الذي افنى حياته من أجل تعليمي الى أعز إنسان ابي العزيز (صالح طبشي)

حفظه الله

الى جدتي وجدي رحمهما الله واسكنهما سفيح جنانه

الى من هم دعمي وسندي في هذه الدنيا , الى من لولا وجودهم لا طعم للحياة اخوتي

والى جميع الأصدقاء دون تفضيل .

كما لا أنسى أن أتقدم بالشكر و الاحترام و التقدير إلى كل الأساتذة بدون استثناء , و إلى كل من ساعدنا لإنجاز هذا العمل من قريب أو بعيد وخاصة الأستاذ "عبدا لمجيد يوسف " لأشرافه على هذا البحث و حرصه و صبره و متابعة لكل معلومة تسجل فيه و على تقديره و احترامه لنا.

والى كافة دفعة تخصص أنظمة اتصالات 2021

طبشي عبير

الإهداء

الحمد لله حمدا كثيرا طيبا مباركا فيه , اهدي تخرجي إلى القلب الحنون و إلى الشفاه التي أكثرت لي الدعاء كلما نطقت إلى التي حملتني وهنا على وهن و سهرت الليالي على راحتي والدتي العزيزة دوغة رقية و إلى خالاتي و أخوالي و كما أن الفضل لوالدي رحمة الله عليه الذي علمني الصعود وهو يراقبني و يشجعني على أن أصل إلى هذا المبتغى و الحمد الله وصلت .

إلى روافد الوفاء إلى نبع الحنان و المحبة أغلى ما أملك إخوتي و أبنائهم (بلخير – عبد الرحمان – بشير الذي بمثابة الأب و الدرع الأيمن – إبراهيم و صاحب الابتسامة الدائمة يعقوب و صاحب القلب الطيب و الرءوف و الوجه البشوش علي و المذلل و محبوب العائلة صاحب القلب الطيب و الأخلاق السامية صلاح الدين اسم)

و إلى المؤنسات الغاليات و عائلاتهم (حاجة , سعيدة , زهية , فاطمة , الزهرة) .

إلى قرة عيني ورفيق دربي و نور حياتي زوجي القدير (بلال عماني) وولديه (عماني عمر و جلولي فاطمة) , و إلى إخوته (آمال , صليحة , فايزة , فتيحة و خالته عائشة و أخوه محمد الشريف و زوجته صبرينة و أبنهما المدلل عبد الرحمان)

إلى فلذة كبدي و قرة عيني ابني حبيبي " محمد فراس عماني " حفظه الله ورعاه و نور دربه و أمه الثانية (فضيلة عماني)

إلى روح أمي التي لم تلدني خالتي (عائشة دوغة) ربي يرحمها كما ربتني صغيرا و أسترتها (الحاج قدور , فاطمة , يمينة , نسيمة , سليمان و ياسين) .

إلى زميلاتي و صديقاتي و كل من أحببته و العائلة

فنتيز أمنة

Abstract

Recently, people are interested in using digital images which leads to a huge growth in their number. The increasing of digital images yields a difficulty to finding and retrieving relevant images. There is a great need to develop an efficient technique for finding the relevant images and understanding user's query. Users of the image search engine usually do not formulate the query in the most efficient manner, possibly because the database does not include user's needs. Most technique failed to retrieve the user's needs or, it retrieves insufficient results. In order to increase the quality of user search results, we expanded the concept used as query exploiting the semantic similarity between the concepts related to the concept query. The main idea is to add new relevant concepts to the query using expansion techniques. By expanding a query by adding new concepts of a user entered term; the recall is increased. To evaluate our work, we compared the results with previous algorithms. The results have shown that query expansion based on semantic similarity improve the recall and retrieves more relevant images compared with the previous algorithms.

Keywords: Image Retrieval, Query Expansion, Semantic Similarity, Ontology.

ملخص

في الآونة الأخيرة ، يهتم الناس باستخدام الصور الرقمية مما يؤدي إلى زيادة هائلة في عددها. تؤدي زيادة الصور الرقمية إلى صعوبة العثور على الصور ذات الصلة واسترجاعها. هناك حاجة كبيرة لتطوير تقنية فعالة للعثور على الصور ذات الصلة وفهم استعلام المستخدم. من المفترض أن المستخدمين لا يقومون دائمًا بصياغة الاستعلامات باستخدام أفضل المصطلحات ، وقد يكون ذلك أيضًا في حالة عدم احتواء قاعدة البيانات على احتياجات المستخدم. فشلت معظم التقنيات في استرداد احتياجات المستخدم أو أنها استردت نتائج غير كافية. من أجل زيادة جودة نتائج بحث المستخدم ، قمنا بتوسيع المفهوم المستخدم كاستعلام يستغل التشابه الدلالي بين المفاهيم المتعلقة باستعلام المفهوم. الفكرة الرئيسية هي إضافة مفاهيم جديدة ذات صلة إلى الاستعلام باستخدام تقنيات التوسيع. من خلال توسيع استعلام عن طريق إضافة مفاهيم جديدة للمصطلح الذي أدخله المستخدم ؛ لقد اختبرنا الخوارزمية الخاصة بنا وقارنا النتائج مع الخوارزميات التقليدية ، وأظهرت التجارب أن توسيع الاستعلام الذي يأخذ في الاعتبار التشابه الدلالي للمفاهيم يعمل على تحسين التذكر واستعادة المزيد من الصور ذات الصلة.

الكلمات المفتاحية: استرجاع الصور ، توسيع الاستعلام ، التشابه الدلالي ، علم الوجود.

Résumé

Récemment, les gens se sont intéressés à l'utilisation d'images numériques, ce qui entraîne une énorme croissance de leur nombre. L'augmentation des images numériques rend difficile la recherche et la récupération d'images pertinentes. Il y a un grand besoin de développer une technique efficace pour trouver les images pertinentes et comprendre la requête de l'utilisateur. Il est supposé que les utilisateurs ne formulent pas toujours les requêtes en utilisant les meilleurs termes, même au cas où la base de données ne contient pas les besoins des utilisateurs. La plupart des techniques n'ont pas réussi à récupérer les besoins de l'utilisateur ou à récupérer des résultats insuffisants. Afin d'augmenter la qualité des résultats de recherche des utilisateurs, nous avons élargi le concept utilisé comme requête en exploitant la similarité sémantique entre les concepts liés à la requête de concept. L'idée principale est d'ajouter de nouveaux concepts pertinents à la requête en utilisant des techniques d'expansion. En développant une requête en ajoutant de nouveaux concepts d'un terme saisi par l'utilisateur ; le rappel est augmenté. Nous avons testé notre algorithme et comparé les résultats avec les algorithmes conventionnels, les expériences montrent que l'expansion des requêtes qui prend en compte la similitude sémantique des concepts améliore le rappel et récupère des images plus pertinentes.

Mots-clés : récupération d'images, expansion de requête, similarité sémantique, ontologie.

Table of Contents

Chapter II : Semantic Approaches for Query Expansion

Chapter 3 :Experimentation and Validation

Experimental setup

List of Figures

List of Table

List d'abréviation

IR : Image Retrieval.

TBIR : Text Based Image Retrieval.

CBIR: Content Based Image Retrieval.

QBIC : Query By Image Content.

QBSE : Query By Semantic Example .

QBVE : Query By Visual Example .

QBT: Query By Text.

QE: Query Expansion.

SS: Semantic Similarity.

MOC : Manual Query Expansion.

IQE : Interactive Query Expansion.

AQE: Automatic Query Expansion.

HQE: Hybrid Query Expansion.

IC: Information Content .

GENERAL INTRODUCTION

1-General Introduction

Due to the rapid growth of the World Wide Web (www) , image retrieval systemsare becoming increasingly important, Social media sites like Flickr, Facebook, and Picasa, which allow users to create, upload and annotate images, are increasingly used these days.Users upload tosocial media, whichareaccompaniedbyinformation such as annotations and comments. This metadata helps in sharing and organizing multimedia content, as well as retrieving and managing images.Since image databases contain a lot of information such as text, image features, users and categories, retrieving images from databases is a difficult task. However, with the increasing use of the Internet, there is a need to develop reliable and efficient methods for image recovery. A new information retrieval technology has emerged capable of identifying similarities between documents containing conceptually similar terms, known as semantic similarity. The latter contains three methods that are used extensively for most applications of information and image retrieval systems, namely: Text-Based Retrieval (TBR), DocumentRetrievalTocontent (CBR), and hybrid approaches to image information retrieval systems.

Semantic similarity between concepts is atoolfor determining thesimilar,or semantic distance between two words according to a specific ontology such as WordNet ontology. In other words, semantic similarity is used to classify concepts with similar "characteristics", semantic similarity is calculated by assigning ontology terms and analyzing their relationships in that ontology.

Semantic similarity seeks to provide effective tools for standardizing information content and distribution through interacting information sources.This has long been recognized as a central problem in the semantic web where related sources need to relate and relay information to one another. The semantic web will also enable users to retrieve the information in a natural and intuitive way, but the search engine does not give the user all the information and the results are sometimes insufficient and do not achieve self-sufficiency.

We are trying to solve this problem by expanding the results, sow searched methods for calculating semantic similarity between natural language terms .

In this work we seek to improve and extend results using query expansion, and query expansion using semantic similarity has promising performance improvements over traditional retrieval methods and retrieve more related images.

Problem :

Retrieving images from a large database is a difficult task because image databases contain a lot of information such as text, image features, users and categories, and due to the increasing use of the databasesin web, there is a need to create reliable and efficient ways to retrieve images from large image databases, and it may be a new way to recover the information has emerged able to detect similarities between documents that contain conceptually similar terms known as semantic similarity.

Semantic similarity seeks to provide effective tools for standardizing and distributing information content through interacting information sources.

This has always been recognized as a central problem in the Semantic Web where related sources need to link and transmit information to each other. Whereas, the semantic web has enabled users to retrieve information in a natural and intuitive way, but the search engine does not give the user all the information and results are sometimes insufficient, not self-sufficient and sometimes unsatisfactory.

The web is the most valuable source of information because it contains documents, information and various sources that can be accessed through traditional search engines. However, organizing this information and documents in a way that facilitates the search and access process is very difficult. In addition, with the constant increase in the volume of information published on the web, it is becoming more difficult for search engines to find the right information.

The Proposed Solution:

We try to solve this problem by extending the results in order to improve the final result during the image retrieval process, and by implanting search methods to calculate the semantic similarity between natural language terms.

We seek to expand the query using semantic similarity.In this work, the Query Expansion (QE) plays an important role in improving web searches and image retrieval. Hence,improving the effectiveness of information retrieval (images), this leads tosatisfy users and to provide them with all the information he/she needs. Expanding the query using semantic similarity has promising performance improvements compared to traditional rewriting techniques**.**

Organize This Message

Continue of the thesis is organized as follows:

In the first chapter, we present an overview of the image retrieval system and its applications, and then we touched upon the formulation of the query in retrieval of images of all kinds.

In the second chapter, we talk about semantic methods of query expansion, where we touched on query expansion, SS semantic similarity, and semantic similarity in image retrieval (first and second level and algorithm) .

In chapter three, we talk about all the steps of designing and implementing of our image search engine, then we describe the search engine and explain our search model operating in the ontology field to find more images relevant to the user's needs.

Finally, we conclude our work, with a general conclusion in which we present the main points of this work and some points of view that may result.

Chapter I :

General on Image Retrieval

1-Introduction:

Nowadays, the search of image is a necessity in all sectors of activity: industrial, legal, medical, scientific, economic, and of course, information technology. Therefore, We must find the information wherever it is and quickly so as not to waste time.Thanks to the Internet we find an array of information.However, it is imperative to sort it and have powerful and easy-to-use search tools to retrieve information and images from a large database of digital images.

The purpose of automatic sorting (indexing) systems is to allow the user to find, in the databases, all the images that are similar to the image in the query. Indexing software is designed as a system that takes a reference image as input and returns the standard of similarity between the reference image and all images in the database. This allows these images to be sorted from most similar to least similar.

There are two main methods : one uses manual text annotations and the other uses descriptors extracted automatically from images.
The first method based on manual textual annotations of the images it is most used today.However, indexing these images takes time, especially with the increasing size of image databases.We also notice many indexing issues related to the fact that the text does not always match the image.

To overcome the defeats of image search systems based on manual text annotations, an image search system based mostly on visual content of image is suggested.

The topic of image search has become a very active topic in the international community for more than ten years. In this chapter we will talk about Image Retrieval (IR)techniques, after that we will explain the applications of IRand their different types.Then we present the formulation of the query in images retrieval (semantic example and visual example) .

2-Image Retrieval:

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. The purpose of image retrieval is to storeand retrieve an image or image sequence related to a query [1]. There are a variety of fields such as information retrieval, computer graphics, database management, and user behavior [2].

Manual image annotations are time consuming, tedious and expensive; to counter this, a lot of research is done on image annotations automatically. Additionally, the advent of social media apps and the Semantic Web has led to the development of a number of online photo annotation tools[3].

In the 1990s, Banireddy Prasaad, Amar Gupta, Hoo-min Toong,[4] and Stuart Madnick developed MIT's first image database retrieval system.

Users from all of these fields have different image demands. Journalists may need photographs of specific events , Designers may order materials of certain colors or shapes. While engineers may request drawings for certain models. Thus, the image retrieval system should facilitate for all these users to locate the images that satisfy their demands through inquiries.

In generally, two different methods have been applied to allow searching of image groups: one relies on retrieving the textual images of images based on the text(Text-based Image Retrieval) and the other relies on retrieving image content information for images based on the content (Content-based Image Retrieval)

There is another method in image processing that can help greatly in meeting user requirements, and this method relies on combining existing textual and visual features to produce a better result. It's called the HYBRID APPROACH

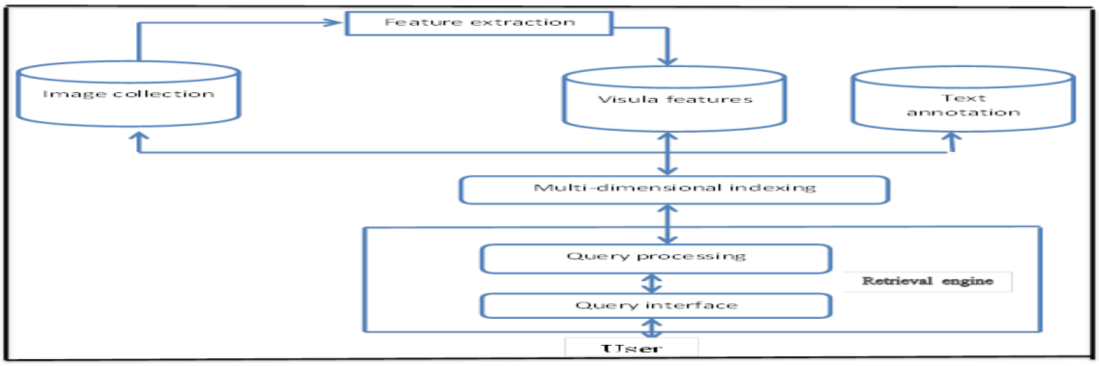


Figure 1: Architecture of image Retrieval System.

2-1-Applications of Image Retrieval:

Images play an important role in transmitting information and are the primary source of content over the Internet.With the rapid advances in information technology and the wide availability of image capture devices such as digital cameras has led to a rapid increase in their number Images.In order to build a robust system that manages and retrieves large image databases accurately and to provide information efficiently.We desperately need technologies that enable us to access, retrieve and process the vast amount of information in these collections.

Image processing is one of the most active areas of research, and researchers have suggested effective methods for us to recover our huge image databases. Digital image processing applications have been used constantly in all different fields of science:

- In the field of medicine, such as: MRI cancer detector.
- Optical character recognition, such as automatic license plate detection .
- Defense
- Facial recognition eg by INTERPOL or Europol to search for criminals i.e. crime prevention.
- Digital libraries.
- Historical research, etc..

2-1-1-Text-based Image Retrieval:

TBIR can be traced back to the late 1970s[5] , the images are indexed using keywords, subject titles, or classification codes, which in turn are used as keys during search and retrieval.

TBIR is currently used in most web image retrieval systems. The text-based approach is keyword based search. This method uses text associated with a picture to perform image retrieval from the image database.

Text based image retrieval systems use methods that vary from a simple iteration method [6]to an ontology based approach. It is assumed that text-based image retrieval systems handle semantic queries more effectively than content-based image retrieval systems.

Early text-based image retrieval systems relied on manual image annotations, with users annotating each image with one or more keywords describing the image's contents. Then these annotations are used to retrieve theimage.

The annotation is time consuming and subjective. Two people can refer to the same image with different annotations. The same person can also tag the same image with different annotations at different times according to the environment. And so the manual explanation

it is used in the context of small fields only, such as personal album, digital library, virtual museum, etc. Chow and Grosky[7] pointed out two problems in hand annotations (1) Synonyms mean that there is more than one way to denote the sameTheme. This leads to a weak recall. (2) Polysemy means that the same word can be used to refer to more than one object. This results in poor accuracy. Hence, we can rely on automatic indexing of images.

There are several methods for automatic indexing. One of them is the calculation of the frequency of occurrence of words. this simple approach can be extended by giving more importance to the words that appear in the image's alt or src tag (the src attribute is required and contains the path of the image you want to include, the alt attribute contains a text description of the image, which is not mandatory but is incredibly useful for the ability to access), or may occur inside the header tag or any other important tags in an HTML document (The HTML <input>src Attribute is used to specify the URL of the image to be used as a submit Button. ... Attribute Values: It contains a single value URL which specifies the link of source image) , Also.

Weights can be given depending on the physical distance of words from the image in the web page. User comments about the resulting results can be used. To improve keyword weights. Decisions about the procedure for hiring eight keywords are also affected by the industry. Finally, these weights considered to retrieve images.

Some text-based image retrieval systems make use of ontology to import outside knowledge of a specific area of interest. Ontology denotes descriptions of the various concepts and the relationships between them[8] . These descriptions areRelationship information is used to retrieve images. Such systems can handle semantic queries more aggressively.

The next figure represents image retrieval based on text.

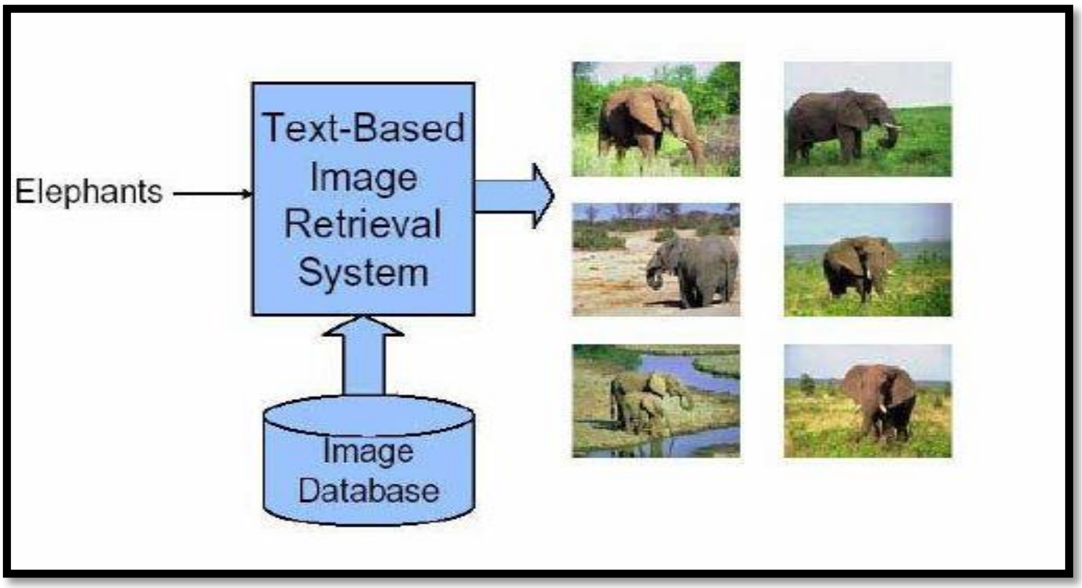


Figure 2: Represents Image retrieval based on text .

Types of text Image Retrieval:

Specialists have divided text image retrieval systems into three categories [9]:

a. Method based on Frequency of Occurrence

This method is the simplest. As the name suggests, this method takes into account the occurrence of keywords. Its keywords The maximum occurrence of a given document is assumed to be important in this document and therefore important from the point of view of the images in that document.

Thus, given the "k" query keyword, the documents are arranged in descending order for the occurrence of the "k" keyword and the images are retrieved from the frequently requested documents as a result.

b. Field Based Method

Domain method "The field-based method (also called the attribute-based and feature-based method) describes an image with one or more field value pairs [8]." This approach generally uses manual image annotations. Initially, the useful features of the images are identified, which collectively describe the image. For each attribute, some indication is given as to what type of value can be assigned. The step for selecting the theme depends on the field. Then in the annotation step, each user of the image (called 18annotator) selects or inputs values for all attributes suitable for the given image.

During retrieval, users are given a simple keyword based query interface, where they can write the query or they can be given the same interface as the annotation interface, where they can specify or enter values of the image attributes. In the other case, users don't need to enter values for everyone. However, entering values for fewer attributes will result in too many images. On the other hand, it takes a long time to enter values for all attributes. A simple way to implement field-based search is to give values to a small number of attributes and start the search..

c. Structure-Based Method

The hierarchy-based approach enables more complex descriptions that include relationships. The structure based method can be considered as an extension of the field based model.

Content based image retrieval (CBIR) came into picture toovercome these problems faced in Text-based image retrieval. InCBIR system images are searched based on their visual contentssuch as color, shape, and texture. [11]

2-1-2- Content Based Image Retrieval (CBIR):

In late 1990's ,content-based image retrieval was introduced by T. Kato. It has been used as an alternative to text based image retrieval. IBM was the first to take the lead by proposing Query by Image Content (QBIC) [12].

CBIR involves the following four parts in system realization:

data collection, build up feature database, search in the database, arrange the order and results of the retrieval.

●The general principle of searching for images:

By content (for image) consists of two stages. During the first phase of offline mode (indexing step), image features are counted and stored in a database. The second stage, known as research, takes place online. User submits a photo as a request. The system calculates the signature in the same way as in the first stage of indexing. Hence, this signature is compared to the set of previously stored signatures in order to return images that are most similar to the order[[13].

During the indexing phase, the signature calculation consists of extracting the visual properties from the images such as:

- **Weaving** (gabor filter, shift to separate waves, etc.)
- **Color** (color histogram, RGB space histograms, TSV, etc.),
- **Forms** (Fourier descriptors, etc.),

After extracting these characteristics, we can compare the images by calculating the distance between them and determining the general similarity scale between them. By measuring this similarity and the desired image.

To find other images as shown in Figure 3 , this type of system does not always need a query image. For example, you can request a search for all images that are somewhat blue, or you can draw a shape and request a search for all images that are close to it.

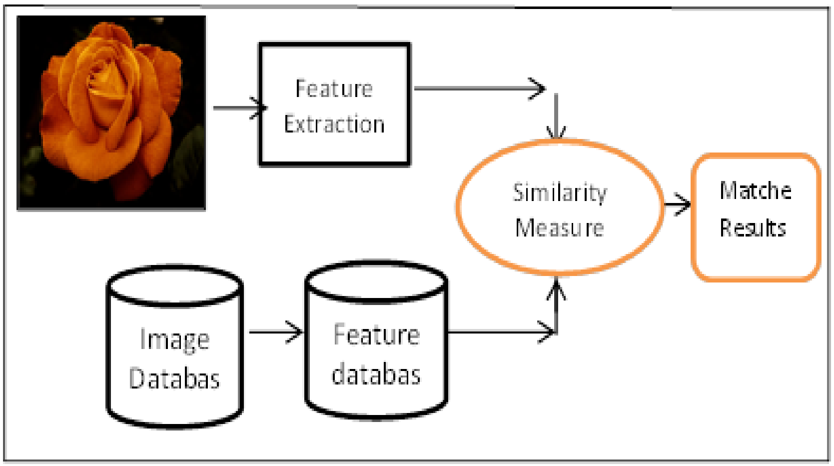


Figure 3: Content-based Image Retrieval System.

◊ **Advantages :**

✓ The features employed by the image retrieval systems include color, texture, shape and spatial are retrieve automatically.

✓ Similarities of images are based on the distances between features.

2-1-2-1-Types of CBIR based Image Retrieval :

a. Region-based : The netra and blobworld are two earlier region based image retrieval systems[14]. During retrieval, auser is provided with segmented regions of the query image, and is required to assign several properties, such as the regions to be matched, the features of the regions, and even the weights of different features [15].

b. Object-based:Object-based image retrieval systems obtain images from a database based on the presence of physical objects in those images. These could be cats, stop signs, helicopters, houses, ears, or any other item that the user wishes to locate.

Also, one of the popular ways to find objects in images is to divide the image in the database first and then compare each segmented region with a region in some of the query images provided by the user.

c. Example-based: In general, image retrieval systems are effective for elements that are easily separated from the background and that have distinct colors or textures [16].

Users provide a reference image, or a section of an image, which the system uses as a starting point for the search. The system then identifies images that are close to the base image.

d. Feedback-based : The system displays a sample of images to the user and requests a ranking. Using these scores, the machine re-queries and repeats the process until the correct image is found [17].

CBIR entails the following four steps in system implementation:

Data collection, Build up feature , search in the database, arrange the order and deal with the results of the retrieval [18]

Data collection using internet spider that can automatically collect networks to conduct internet interviews and collect images on the website, then pass through all other networks through a URL, and it repeats this process and collects all the images it reviewed on the server. This system is based on indexing.

Feature build We analyze the collected images and then extract the feature information. Currently, the most frequently used features include such as color, texture, shape and other industry-specific features that can be extracted. Features are extracted from the query and other images stored in the database, based on their pixels.

CBIRS stores image information in compressed form in a separate database known as a feature database, also known as image signature.

search in the database The system extract the feature of image that waits for search when user input the image sample that need search, then the search engine will search the suited feature from the database and calculate the similar distance, then find several related webs and images with the minimum similar distance..

Process and index the results the user indexes the image obtained from the search based on the similarities between the features, then the retrieval images are returned to the user and let him choose.

And if the user is not satisfied with the search results, he can retrieve the image and re-search the database.

2-1-2-2-Feature Extraction:

The main features based methods aredescribed as following :

Color: Color is an important visual factor and feature of content based on image retrieval systems. Depending on the different applications, different color spaces are available [19].

Color images are represented using the color space. RGB space, representing the entire color of gray based on a combination of red, blue and green intensity.

Includes a variety of color spaces, RGB, LUV, HSV (HSL),YCrCb and descriptors in CBIR systems include the color contrast matrix, chromatogram, color moments, and color consistency vector [20].

A Color Structure Descriptor (CSD) represents an image by both the local structure of the color and the color distribution of the image or image area.

Shape : Natural objects are recognized primarily by their shape. Some features related to object shape are calculated for each specific object within each stored image. Representations of the figure can be divided into two categories, both on the basis of boundaries (only the outer boundaries of the figure are used) and on the basis of area (the entire area of the figure is used). The most successful representation of these two categories is the use of Fourier descriptor and instant variables [21].

The central idea of the Fourier descriptor is to use the converted Fourier boundaries as a feature of the form. The main idea of moment constants is to use Region-based moments, which are constant for transformations as a feature of the form.

You can enter queries into the system either in the form of image or as a diagram.

Texture: Texture means visual patterns that have the property of homogeneity and cannot be caused by a single color or intensity. Examples of these surfaces include clouds, trees, bricks, hair, and weaving. The concept of texture generally refers to a spatial pattern that has some characteristics of homogeneity [22] . Vectorial features are extracted to capture image texture information. The six visual texture characteristics were roughness, contrast, directivity, streak resemblance, regularity, and roughness.

These characteristics are said to be low-level, because they are very close to the signal, and do not convey any particular semantics in the image. After extracting these characteristics, the comparison generally entails determining different distances between these characteristics and defining a measure of the general similarity between two images by measuring this similarity, and the set of similarity measures for the desired image.

2-1-2-3-Applications of CBIR:

There are various possible applications for CBIR technology has been identified. Some of these are mentioned below:

- Investigations: face recognition systems, copyright on the Internet
- Shapes identification: identification of defect and fault in industrial automation.
- Medical diagnosis: Tumours detection, Improve MRI and CT scan Understand ability.
- Journalism, advertising Media, Fashion and graphic design.
- Remote sensing: Various information systems, weather forecast, satellite images.
- Trademark databases, Art galleries, museums and archaeology.
- Architectural and engineering designs.
- Cartography: map making from photographs, synthesis of weather maps.
- Digital Forensics: finger print matching for crime detection.
- Radar engineering: helps in detection and identification of targets

In general, CBIR can be described in terms of following stages:

a. Identification and utilization of intuitive visual features.

b. Features representation

c. Automatic extraction of features.

d. Efficient indexing over these features.

e. Online extraction of these features from query image.

f. Distance measure calculation to rank images.

Following are some of the CBIR based systems:

1. PicSOM Image Browsing System http://www.cis.hut.fi/picsom
2. QBIC system (IBM) http://www.qbic.almaden.ibm.com/

3.Hybrid Approach

In the proposed approach, linkages between images are used to improve the image retrieval technique. Correlation between images is not taken into consideration upon exit from work [23].

Text and content based approaches have advantages and disadvantages. Portions of the disadvantages of these two methods can be resolved by combining them.

That is, this approach takes a different position , treating images and text as interchangeable data. by estimating the co-distribution of features and words and providing annotations as statistical inference in a graphical model, this mixed approach combines existing textual and visual features to produce a better result [24].

This method aims to discover the relationship between visual features and textual word, and as a result, the combination of a text-based and content-based image retrieval approach is insufficient to deal with the problem of image retrieval in large-scale databases.

There are a lot of applications where you use pictures; Hence, the image retrieval systems will facilitate their work. Among them [25]:

- Education and training
- Travel and Tourism
- Fingerprint identification
- Recognize faces
- Control system

- Home entertainment
- Fashion, architecture and engineering
- Historical and artistic research

4-Query Formulation in Image Retrieval:

4-1-Query By Semantic Example QBSE:

QBSE extends the idea of querying by example in the field of semantic image representations. The semantic vocabulary is identified first, and the semantic retrieval system is trained to name each image with the posterior probability of each concept appearing in the vocabulary.

The starting point for any retrieval system is an image database D ={I1…….ID}

Images are observations from a random variable **X ,** defined on some feature space x . In the absence of labels, each image is considered an observationfrom a different class, determined by a random variable **Y** defined on . In this case, the retrieval system is said to operate at the visual-level. Given a query image **Iq** , the MPE retrieval decision is to assign it to the class of largest posterior probability, i.e.,

$$\boldsymbol{y^{*} = argmax\ P(y/x)(y/Iq\ )}\text{………………………………..}\textbf{(I . 1)}$$

Note that ci is a binary L-dimensional vector such that Ci,j = 1 if the i^{th} image was annotated with the j^{th} keyword in L.

The training set D = {(I1, c1),...,(ID, Cd)} of image-caption pairs is said to be weakly labeled if the absence of a keyword from caption ci does not necessarily mean that the associated concept is not present in Ii.

This is usually the case in practical scenarios, since each image is likely to be annotated with a small caption that only identifies the semantics deemed as most relevant to the labeler.

The design of a QBSE retrieval systems requires two main components. The first is a semantic image labeling system that, given a novel image I, produces a vector of posterior probabilities π = (π1,...,Πl)T for the concepts in L.

This can be seen as a feature transformation, from the space of image measurements X to the L-dimensional probability simplex SL, i.e. a mapping $\Pi : X \rightarrow SL$ such that $\Pi(I) = \pi$.

Each image can, therefore, be seen as a point π in SL, i.e. the probability distribution of a multinomial random variable defined on the space of semantic concepts.

The second component is a query-by-example function on SL.

i.e. $f : SL \rightarrow \{1,...,D\}$ such that

$f(\pi)$ = arg maxi $s(\pi, \pi i)$ where π is the query SMN πi the SMN that

char-acterizes the I the database image, and $s(\cdot, \cdot)$ an appropriate similarity function.

Given that SMNs are probability distributions, a natural similarity function is the Kullback-Leibler divergence [26]

$$\boldsymbol{S(\pi, \pi') = KL(\pi || \pi i) = \sum_{i=1}^{l} \pi i \ \log \frac{\pi' i}{\pi i}} \quad \text{...............................(I . 2)}$$

4-2-Query By Visual Example QBVE:

A QBVE system operates at the visual level and assumes that the feature vectors which compose any I image are sampled independently

$$\boldsymbol{Px/y(\mathrm{I} \mid y) = \prod_J Px/y(Xi \mid Y)} \text{ . ...(I . 3)}$$

Some density estimation [21] procedure is used to estimate the distributions PX/Y(X/y).This produces a vector of parameters Ty per image, e.g.,

$Ty = \{u_y^j, \sum_y^j, \alpha_y^j\}$, y=1,……..,D when

$$\boldsymbol{P(x/y)(\mathrm{X} \mid y; \tau_Y) = \sum_J \alpha_y^j \, \partial(X, \mu_y^j, \Sigma_y^j 0)} \text{ ....(I . 4)}$$

is a mixture of Gaussians. Here, α_Y is a probability mass function such that

$\sum_j \alpha_y^j = 1$, $\partial(X, \mu, \Sigma)$ a Gaussian density of mean μ and covariance Σ, and j an index over the mixture components.

Image retrieval is based on the mapping $g : X \rightarrow \{1, \ldots\ldots D\}$of (1), implemented by combining (2), (3) and Bayes rule. Although any prior class distribution Py(i) can be supported, we assume a uniform distribution in what follows [27].

4-3-Query By Text QBT:

We present an overview of the key approaches from that domain because concept-based query expansion is related to research in text-based query expansion. In theory, the concept is sound is to add new query terms to the original query that are relevant to the search The inclusion of phrases that are relevant can increase recollection by finding related documents through matches, which is especially useful for brief queries to the additional terms. It may also be used to narrow the scope of overly broad queries, re-ranking the results and increasing precision. Of course, this is true Only works if the revised query is consistent with the original. Experiments in text document retrieval, on the other hand, have revealed that query expansion is strongly subject dependent.

Can increase recollection by finding related documents through matches, which is especially useful for brief queries.[28]

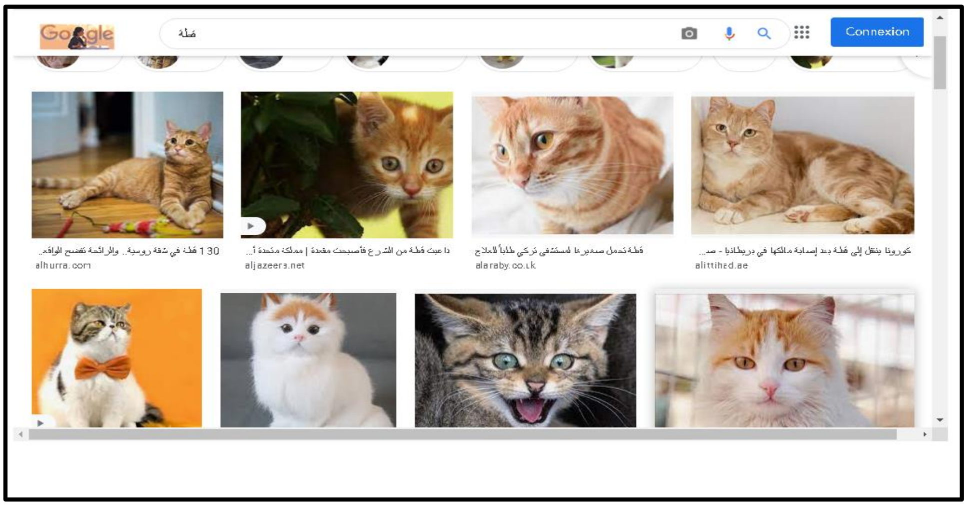


Figure 4: Search results for the word "Cat ".

4-3-1- Methods for extending the query with text QBT:

- **Lexical approaches (language-specific)**

Synonyms and other linguistic word associations are used in lexical techniques to take advantage of global language features (e.g., hypernyms). Typically, these methods are based on on dictionaries or other knowledge representations of a similar nature WordNet [29] is an example of a reliable source. Approaches to lexical query expansion can help with recall, but word sense is still important When there's a lot of ambiguity, it's easy to get off track, as more terms are added to the query, the semantics of the query changes. .

- **Statistical approaches (corpus-specific)**

Statistical techniques are data-driven and aim to find meaningful word associations based on co-occurrence of terms selection of features and analysis these are more generic relationships that may or may not have a language interpretation. Early Words were grouped together using corpus analysis methods based on their co-occurrence patterns in documents [30]. Term clustering [31] and Latent Semantic Analysis are two related approaches based on term-document cooccurrence, indexing [32] groups related terms into clusters or hidden orthogonal dimensions. Later techniques aim to keep the conversation on track .By only looking for frequently co-occurring patterns inside the same context, rather than across the full document, where the context can be the same paragraph, sentence, or simply the same word

- **Statistical approaches (query-specific)**

Local analysis uses only a subset of the documents, as opposed to global statistical approaches that evaluate the distribution and co-occurrence of words throughout a whole corpus to find substantial patterns of co-occurrence this subset is usually a collection of documents that the user has submitted or labeled as relevant to the query. For example, with relevance feedback systems, the system adjusts the query based on users' relevance judgments of the documents obtained [33]. Some systems simply assume that the top N retrieved documents are relevant, where N is decided empirically and is typically between 20 and 100. This eliminates or reduces the requirement for user feedback. This is a result of the presumption that the top results are more relevant than a random selection, and that any strong co-occurrence patterns discovered within this group are more likely to be related to the top results query. This method is known as pseudo-relevance feedback [34] .

5- Conclusion:

Advances in data storage and image acquisition technologies have enabled the creation of large sets of image data, where image search has become a necessity due to the current technological development .

Image search engines fall into two main categories: one that uses manual text annotations and one that uses descriptors automatically extracted from images. Each of these categories has advantages and limitations. Some engines have combined both technologies to improve search quality.

In this chapter, we have tried to provide an overview of the most common approaches for the different types of image retrieval systems. First, we introduced the concept of image retrieval, then we talked about the techniques used in image retrieval and presented a set of constructs aimed at defining the relevant basic concepts. Describe the main issues to consider when designing this type of image retrieval system.

We have outlined the approach and explained the working principle of each approach. Second, we talked about the problem of retrieving images by query by semantic example (QBSE) and via visual example (QBVS).

Image retrieval is a field of computer research that can be used to browse, search, and retrieve images from a large image database.

Chapter II:

Semantic Approaches for Query Expansion

1- Introduction:

Since the 1940s , the problem of Information Retrieval (IR) has attracted increasing attention, especially because of the dramatically growing availability of documents. IR is the process of determining relevant documents from a collection of documents, based on a query presented by the user There are many IR systems based on Boolean, vector, and probabilistic models.

Search engines call on the search query expansion to increase the quality of the user's search results. it is assumed that users don't always formulate search queries using the best terms. Perhaps the best in this case is that the database does not contain the conditions entered by the user .

Query expansion (QE) is the process of reformulating a specific query to improve retrieval performance in information retrieval operations, particularly in the contxt of an understanding query . In the context of search engines, query expansion involves evaluating user input and extending the search query to match additional document . Search query expansion includes techniques such as :

_ Find synonyms for words

_ Find related semantic keywords

_ Find all forms of morphological words by derivation for each word in the search query

To find more accurate and relevant results , it is necessary to reinforce the original query with words that are synonymous or related to te search words , in order to improve the effectiveness of the information retrieval process. One way to extend the query is to use semantic similarity.

Semantic similarity between concepts is a method of measuring semantic similarity, or the semantic distance between two concepts according to a specific ontology. In other words, semantic similarity is used to define concepts that have common "properties". Although man does not know the formal definition of the relationship between concepts, human can judge the relationship between them. For example, a young child might say that an “apple” and “peach” are more related to each other than an “apple” and “tomato”. These concept pairs are related to each other and the definition of their structure is formally called the "is-a" hierarchy. Semantic similarity methods have become extensively used for most applications of semantic information retrieval systems based on intelligent knowledge (determining the optimal match between query terms and documents) and demystifying meaning and bioinformatics .

In this chapter we will introduce the concept of expanding the query and also discuss the semantic similarity and methods of measuring similarity between words and identify one of the cases of information retrieval with semantic similarity

2 - Query Expansion:

Query expansion (QE) is a computer science technique that is studied in the realms of natural language processing and information retrieval.

The method of reformulating a seed query to enhance retrieval is known as query expansion efficiency in operations involving information retrieval. The term "quest" is used in the context of web search engines Expansion is the process of analyzing a user's input (the words typed into the search query field, as well as other types of data) and extending the search query to match more data the records .

Query expansion is the process of introducing new words and phrasesnew images in our case to current search terms in order to create a more comprehensive query. Previous query expansion approaches, on the other hand, were limited to extracting expansion terms from a subset of documents and did not take advantage of the cumulative data on user experiences [35] , there are four different ways to expand the query:[36]

Manual Query Expansion(MQE): is based on the user's dexterous judgment, which involves manually selecting candidate words and reformulating the initial query. Manual labor, on the other hand, has been found to be beneficial in previous studies Just 25% of the related documents in the database can be retrieved by choosing candidate words for the expansion (Sharma, Pamula, & Chauhan, 2019) [37] .

Interactive Query Expansion (IQE) : is also known as semi-automatic, in which the machine suggests terms and the user chooses the appropriate expansion terms to access more applicable documents from the database the next iteration.

Automatic Qery Expansion (AQE) : Each candidate term's weight is calculated, and the highest weighted terms are chosen to be added to the initial query (the algorithm performs the entire process without human intervention).

A hybrid method for query expansion (HQE) refers to combining two or more methods for the expansion process, such as the hybrid approach suggested by Han and Chen (2009), [38], which combines two techniques: neural networks and ontologies.

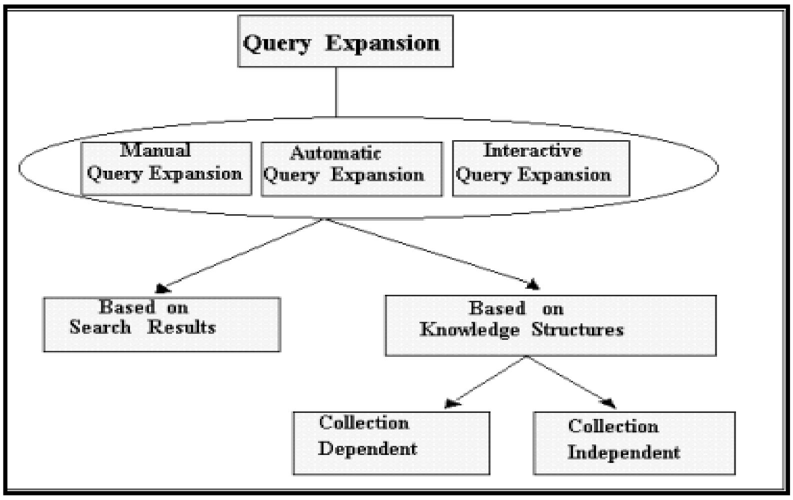


Figure 5: Query Expansion: Methods and Source .

3- Semantic Similarity SS:

Semantic similarity between concepts is a method of measuring semantic similarity, or the semantic distance between two concepts according to ontology.It is also used to identify concepts that have common characteristics. Although a person does not know the formal definition of the relationship between concepts, he can do kinship between them. For example, we can say that “oranges” and “bananas” are more related to each other than “Banana” and “tomato”. These pairs of concepts are related to each other .[39]

Semantic similarity and semantic association are two related words, but semantic similarity is more specific than relatedness . Semantic similarity is a measure defined across a set of documents or terms, where the idea of distance between items is based on similarity in their meaning or semantic content rather than lexical similarity. They are mathematical tools used to estimate the strength of the semantic relationship between language units, concepts, or examples, through numerical description obtained according to a comparison of information that supports its meaning or describes its nature. The term semantic similarity is often confused with semantic correlation. A semantic association involves any relationship between two terms, while a semantic similarity involves only "it" relationships. For example, the word "car" is similar to "bus", but is also related to "road" and "driving" Computationally, semantic similarity can be estimated by determining topological similarity, using ontology to determine the distance between terms/concepts.. [40]

3-1- Chronological study of arc-Based Measurements

The number of arcs that separate two concepts in a taxonomy is counted by similarity measures that are based on arcs (Tchechmedjev, 2012)[41]. In this part, we'll refer to Figure 6 to illustrate the various events. On this diagram, the concepts $c_1$and $c_2$are two conceptions that share a common ancestor,c_3.N_1, N_2, and N_3represent the number of arcs betweenc_3 and c_1, $c_3$and c_2, and $c_3$and the racine . We'll also define the depth of a notion in a taxonomy as the level of that concept in relation to the taxonomy'sracine, it is noted 'Pi.

The total depth of a hierarchical structure is the maximum value of the depths of all these elements. It is noted PD. [42]

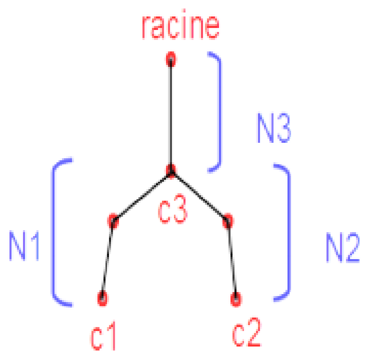


Figure 6: Example of taxonomy for edge-based similarity measures .

This section briefly discusses corpus-based techniques (Section 3.1.1) and knowledge-based semantic similarity metrics that have shown to be effective in NLP and IR applications.

3-1-1- Corpus-Based Approaches:

Corpus-based methods use knowledge from broad corpora like Wikipedia to test semantic similarity between concepts. Some people have adopted this concept Pointwise Mutual is an example of a work that takes advantage of idea connections [45]Information or [44] Normalized Google Distance , while some other works use distributional semantics techniques to represent the concept meanings in high-dimensional vectors such as Latent Semantic Analysis [46] and Explicit Semantic Analysis [47]. Recent works based on distributed semantics techniques consider advanced computational models such as Word2Vec [48] and GLOVE [49], representing the words or concepts with low-dimensional vectors.

The co-occurrence knowledge of words in the same sense will lead to a wide range of words being considered associated. Due to the fact that corpus-based methods they usually calculate the general semantic relatedness between words based on contextual knowledge of words rather than the semantic similarity that exists between words [50]. It is dependent on hierarchical relationships . Furthermore, corpus-based semantic similarity approaches treat concepts as terms without distinguishing between their various meanings (word senses). In contrast to knowledge-based approaches that depend on Corpus-based methods usually have greater vocabulary coverage on KGs due to their computational models can be extended to a variety of corpora, both old and new they are modeled after words because they are built on words. rather than definition taxonomies, and textual corpora.

In the following section, we will look at corpus-based methods and provide a thorough overview of the key knowledge-based methods. [43]

3 -1-2- Knowledge-Based Approaches:

K nowledge-based approaches measure the semantic similarity of concepts in KGs. We first give a formal definition of KG

<u>Definition 1</u>: A KG is defined as a directed labeled graph,$G(V,E,\tau)$

where V is a set of nodes, E is a set of edges connecting those nodes; and $\tau$is a function V×V ! E that defines all triples in G.

Knowledge-based methods use a KG to compare the semantic similarity of concepts

$c_1, c_2 \in V$ formally sim(c_1, c_2) KG's semantic information will be used.

The semantic distance between concepts, which is usually expressed by the path connecting two concepts in KG, is the most intuitive semantic knowledge. On the surface, the more similar two concepts are, the shorter the route between them is.

<u>Definition 2:</u> A path P(c_1, c_2) between c_i, c_j ∈V through G is a sequence of nodes and edges P(c_i, c_j)= $\{c_i, e_i, \ldots\ldots u_k, u_{k+1}, \mathrm{e}_{\mathrm{k+1}}, \mathrm{c}_{\mathrm{j}}\}$connecting the concepts $c_i$and c_j with cardinalityor size n. For everytwo consecutive nodes$u_k u_{k+1} \in V$ in P(c_i, c_j), there exists an edge$e_k \in E$.

note that, despite the fact that KG is modeled as a guided graph, we ignore edge direction since semantic relations can be considered semantically sound inverse relationship [51]

Let Paths $(c_i, c_j) = \{P_1, P_2, \ldots. P_n\}$ be the set of connections between the definitions ci and cj and cardinality or size N .Let $|P_i|$ denote the length of a path $P_i \in Pat□s\ (c_i, c_j)$, then length $(c_i, c_j) = \min_{1\le i\le N}(|\mathrm{P_i}|)$ The shortest path length between two definitions is denoted by this symbol. To find the shortest path length between definitions, the path [47] method is used reflect the semantic distance between them, and the distance may be any length of time transformed into resemblance.[43]

◊ **Measure of Rada**

The emantic similarity measures, which are based on arcs, have been introduced by (Rada et al., 1989) [53]. They have and are described in terms of the distance that separates two definitions. The expression 5 provides the measurement : [42]

$$\textbf{Sim Rada } (\boldsymbol{c_1}, \boldsymbol{c_2}) = \frac{\mathbf{1}}{\mathbf{1+dist(c_1,c_2)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textbf{(II . 5)}$$

$$= \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{N_1} + \boldsymbol{N_2}}$$

The lch [54] method uses a non-linear function to quantify semantic similarity between concepts based on their shortest path length **.**

◊ . Measure of Leacock & Chodorow

Leacock & Chodorow (Leacock et Chodorow, 1998) [55] were inspired by the work of (Rada et al., 1989)[53] and (Resnik, 1999)[64] in order to define a semantic similarity measure. This measurement is defined by the phrase 6 : [42]

$$\boldsymbol{sim_{LC}}\,(\boldsymbol{c_1}, \boldsymbol{c_2}) = -\,\mathbf{log}\Big(\frac{\boldsymbol{dist}\,(\boldsymbol{c_1}\,, \boldsymbol{c_2})}{\mathbf{2} \times\, \boldsymbol{P_D}}\Big) \quad \dots\dots\dots\dots\dots..\dots..(\ \mathbf{II.6})$$

where D is the idea taxonomy's maximum depth in a KG. Depth refers to the path taken by hierarchical relations between the root idea and a given concept due to the fact that KGs contain concepts that can be organized as a taxonomy of concepts having hierarchical relationships, such as to name a few: WordNet taxonomy, DBpedia ontology class a few of.

<u>**Definition 3:**</u> The $dept\square(c_i) = lengt\square(c_i, c_{root})$ of a concept$c_i \in V$is defined as the shortest path length from c_i to root concept $c_{root} \in V$. For every two consecutive nodes $u_k, u_{k+1} \in P\ (c_i\ , c_{root}\ )$, there exists an edge$e_k \in \{\square ypernym\ , subClassOf\}$

The definition of using depth knowledge of concepts to calculate semantic similarity stems from the fact that higher-level concepts in a taxonomy are meant to be more general. As a result, the similarity between lower-level concepts should be considered more comparable than the similarity between higher-level concepts concepts at a higher level In Figure 6, for example, the definition The concept of a scientist and an actor is more similar to the concept of a scientist and an actor combine an individual and a product .

The Least Common Subsumer (LCS) is the most basic term in the two concepts' mutual ancestor. The LCS of concept scientist and concept actor, for example, is the person with a definition Let clcs stand for the LCS of concepts ci and cj, respectively the wup [56] method uses the following to determine semantic similarity between two concepts.[43]

◊ **Measure of Wu & Palmer**

(Wu et Palmer, 1994) [57] described a knowledge-based approach to the problem

In order to create a KBMT (Knowledge Based Machine Translation) machine, Chinese translation of the English language. By clicking on Figure 1, you can see the degree of similarity can be expressed in a number of ways 7: [42]

$$\mathbf{sim}\ (\mathbf{c_1}, \mathbf{c_2}) = \frac{\mathbf{2 \times P_3}}{\mathbf{P_1 + P_2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathbf{II\ .7})$$

$$= \frac{\mathbf{2} \times \boldsymbol{N_3}}{\mathbf{N_1 + N_2 + 2 \times N_3}}$$

◊**Measure of Sussna**

(Sussna, 1993)[58] proposes an emantic measure of proximity that is based on the ponderation of concepts. The notion of proximit es emantique is broader than that of similarite s emantique (Sy, 2012), car elle ne se limite pas ´ a la relation de subsomption ` (la relation de hierarchie dans une taxonomie not'ee is-a)

The distance between two adjacent nuds in a taxonomy is determined by the weight of the relationships that connect them. She is given by the second expression 8 : [42]

$$\boldsymbol{ww(c_x,c_y)} = \frac{\boldsymbol{w(c_x \to R\ c_y) + w(c_y \to R^{-1}\ c_x)}}{\mathbf{2\ \times max(p_x, p_y\ )}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\boldsymbol{II}\ \mathbf{.8})$$

O $w(c_x \to R\ c_y)$ et $w(c_y \to R^{-1}\ c_x)$ represent the weight of the relation R and the weight of the inverse R^{-1} between the two concepts, respectively. The relationship's weight is determined by the expression 9:

$$\boldsymbol{w(c_x \to R\ c_y)} = \mathbf{max}\ \boldsymbol{R} - \frac{\boldsymbol{max\ R - min\ R}}{\boldsymbol{n_R(c_x)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\boldsymbol{II}\ \mathbf{.9})$$

avec max_R et min_R qui representent respectivement la valeur maximale et la valeur ´ minimale que nous pouvons associer a une relation ` R ; n_R (c_x) represente le nombre ´ d'arcs de$c_x \ \to^R \ c_z$. [42]

◊ **Measure of Hirst & St Onge**

(Hirst et St-Onge, 1998)[59] described a measure of semantic proximity between two concepts (classes) that takes into account changes in direction. Two definitions are semantically related if they are linked by a short path (five arcs at most maximum) et que les direction changes ne soient pas excessively frequent .

When two separate directions are intersected in the same route, we refer to it as a direction change (Hirst et St-Onge, 1998). As a result, the measure is described by the relation 10 : [42]

$$\boldsymbol{Rel_{HSO}(c_1, c_2) = C - diste(c_1, c_2) - K \times virages(\ c_1, c_2)} \ldots\ldots\ldots\ldots\ldots \quad \textbf{(II . 10)}$$

$$\boldsymbol{= C - (N_1, N_2) - K \times virages\ (c_1, c_2)}$$

With C and k which are two constants (C = 8 and k = 1); turns (c_1 ,c_2) which indicates the number of changes of direction .

◊ **Measure of Stojanovic**

(Stojanovic et al., 2001)[60] uses the depth of concepts in a hierarchical structure to evaluate their semantic similarity. This measure emphasizes a generalized version of the definition of depth in order to account for multiple heirs. A sa the expression 11 provides the following formula: [42]

$$\boldsymbol{sim_{sto}(c_1\ , c_2) = \frac{P_3\ \ + 1}{(P_1 + 1) + (P_2 + 1) - (P_3\ + 1)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots. \quad \boldsymbol{(II\ .11)}$$

◊ **Measure of Zhong**

The measure of (Zhong et al., 2002**)**[61] assesses the lack of conceptual similarity. The proposed approach is based on the distance between definitions. The similarity measure is expressed as follows: [42]

$$\boldsymbol{sim_{zhong}(c_i\ , c_j) = 1 - \mathrm{dist}\ (c_i, c_j)} \ldots\ldots\ldots\ldots\ldots\ldots..\boldsymbol{(II\ .12)}$$

The distance is determined by the expression 13 :

$$\boldsymbol{dist(c_i, c_j) = \frac{1}{2^{p_{pppc}}} - \frac{1}{2^{p_i+1}} - \frac{1}{2^{p_{j+1}}}} \ldots\ldots\ldots\ldots\ldots\ldots. (\boldsymbol{II.13})$$

C_{pppc} designe le definition commun ement appel e le plus petit parent commun de c_i etc_j avecC_{pppc} .P_{pppc} , p_i , and p_j reflect the depths ofC_{pppc} , c_i , andc_j , respectively

◊ **Measure of Zargayouna**

(Zargayouna (2004**)**[62] proposes an expansion of Wu & Palmer's measure by taking into account the lowest level of taxonomy, which he refers to as the bottom. He adds a measure of specificity to Wu & Palmer's measure, which considers the degree of specificity of the concept. In other words, it is the number of arcs that separates him from the rest of the world at the heart This measurement is expressed using the formula 14 : [42]

$$\begin{cases} \boldsymbol{sim\,(c_1, c_2) = \dfrac{2 \times P_3}{P_1 + P_2 + spec(c_1, c_2)}} \\ \boldsymbol{spec(c_1, c_2) = P_b(C) \times dist(c_1, c_3) \times dist(c_2, C)} \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\boldsymbol{II.14})$$

$P_b(C)$ is the number of arcs that separate c_3 from bottom (c_3 represents the c_{pppc})

◊ **Measure of Zhong**

A similarity measure based on Wu and Palmer's measure is also proposed (Slimani et al., 2007) [63]. This measure was implemented to address some of Wu and Palmer's measurement problems related to their hierarchical structure The measure proposed by (Slimani et al., 2007) is given by the formula 15 : [42]

$$\boldsymbol{sim_{tbk}(c_1, c_2) = \frac{2 \times P_3}{P_1 + P_2} \times f_p(c_1, c_2)} \ldots\ldots\ldots\ldots\ldots\ldots (\boldsymbol{II.15})$$

With

$$\boldsymbol{f_p(c_1, c_2) = \begin{cases} \dfrac{1}{|P_1 - P_2| + 1} & Si\ c_1\ and\ c_2\ are\ included\ in\ the\ same\ path \\ 1\ if\ not \end{cases}} ; \ldots.. (\boldsymbol{II.16})$$

◊**Measure Resnik**

The measure introduced by Resnik (Resnik, 1995) [64]returns the information content (IC) of the LCS of two concepts: [65]

$$sim_{res} = IC\ (LCS) \ \ldots\ldots\ldots\ldots \quad (II\,.17)$$

where IC is defined as:

$$IC_{(c)} = -\log P_{(c)} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (II\,.18)$$

and P(c) is the probability of encountering an instance of concept c in a large c

◊**Measure Lin**

The metric introduced by Lin (Lin, 1998) [66], which builds on Resnik's measure of similarity, and adds a normalization factor consisting of the information content of the two input concepts [65]

$$sim_{lin} = \frac{2 \times IC\ (LCS\ )}{IC_{(c1)} + IC_{(c2)}} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (II\,.19)$$

◊**Measure Jiang & Conrath**

Finally, the last similarity metric we consider is Jiang & Conrath **(**Jiang and Conrath, 1997) [67], which returns a score determined by: [65]

$$sim_{jnc} = \frac{1}{IC_{(c1)} + IC_{(c2)} - 2 \times IC(LCS)} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (\mathrm{II}\,.20)$$

3-1-3-Advantages and Disadvantages:

In this section, we present Table 1, which summarizes our chronologic study and lists the benefits and drawbacks of similarity measures.[42]

Measures	Year	Advantages	disadvantages
Rada	1989	Simple and easy to execute	No consideration is given to the depth of concepts; no adaptation to WordNet is made
Sussna	1993	Take note of all the taxonomic relationships	Very expensive in terms of calculation
Wu & Palmer	1994	Simple and easy to implement; considers the depth of the data conception	There isn't much of a resemblance between concepts like sin and concepts from the same hierarchical level
Leacock & Chodorow	1998	Simple a implemente	Only takes into account the is-a relation; less efficient as Wu & Palmer on Word Net
Hirst & ST Onge	1998	Allows you to compare the similarity of a name and a verb on WordNet	There are limits on the number of paths available
Stojanovic	2001	Take into account numerous ancestors; boost Wu & Palmer on WordNet	It does not produce a strong resemblance between neighboring concepts and concepts of the same kind hierarchical system
Zhong	2002	Gives a better similarity between "father" and "son"` " that between two "brothers" in a taxonomy	No guarantee of uniqueness of the smallest common parent (pppc)
Zargayouna	2004	Simple and easy to bto implement ; takes count the depth of concepts and the similarity between related concepts and concepts of the same hierarchy	Too dependent on the organization of concepts in the Taxonomy
Slimani	2007	Simple and easy to to implement ; take in count the depth of concepts and the similaity between related concepts and concepts of the same hierarchy	Too dependent on the organizatio sation of concepts in Taxonomy

Table 1: Advantages and disadvantages of similarity scales .

4- Semantic Similarity in Image Retrieval :

In this mode of search, the user may formulate his or her query using concepts , as well as ontological relationships. The system then conducts a search based on these criteria concepts for traversing the entire set of corresponding metadata in order to find photos with these metadata annotated . and finally the result of this search is displayed on behalf of the user. This system consists of two levels namely: [68]

5- Query processing steps :

1- The user expresses his needs using images that express his request.

2- Processing is done automatically as related words and synonyms are imported from manually created dictionaries for image indexing.

3- ore documents are matched and the percentage of similarity between them is studied. Alternative word models are also matched to the term entered by the user, which leads to increased retrieval and the inclusion of the most relevant result set pages.

4- Expand the query, reorder and expand the request to achieve the purpose of improving the retrieval rate.

5- It presents documents and concepts that are closely related to the user's request, in which the properties and relationships associated with the information resources in the database are queried.

6- Displaying the results to the user according to his request, as it tends to provide documents with a high density (high frequency) at the top of the search results, where these results are more relevant and closely related to the user's question. Which leads to the quality of the results despite the greater recovery.

The stages of query processing can be described in the following diagram:

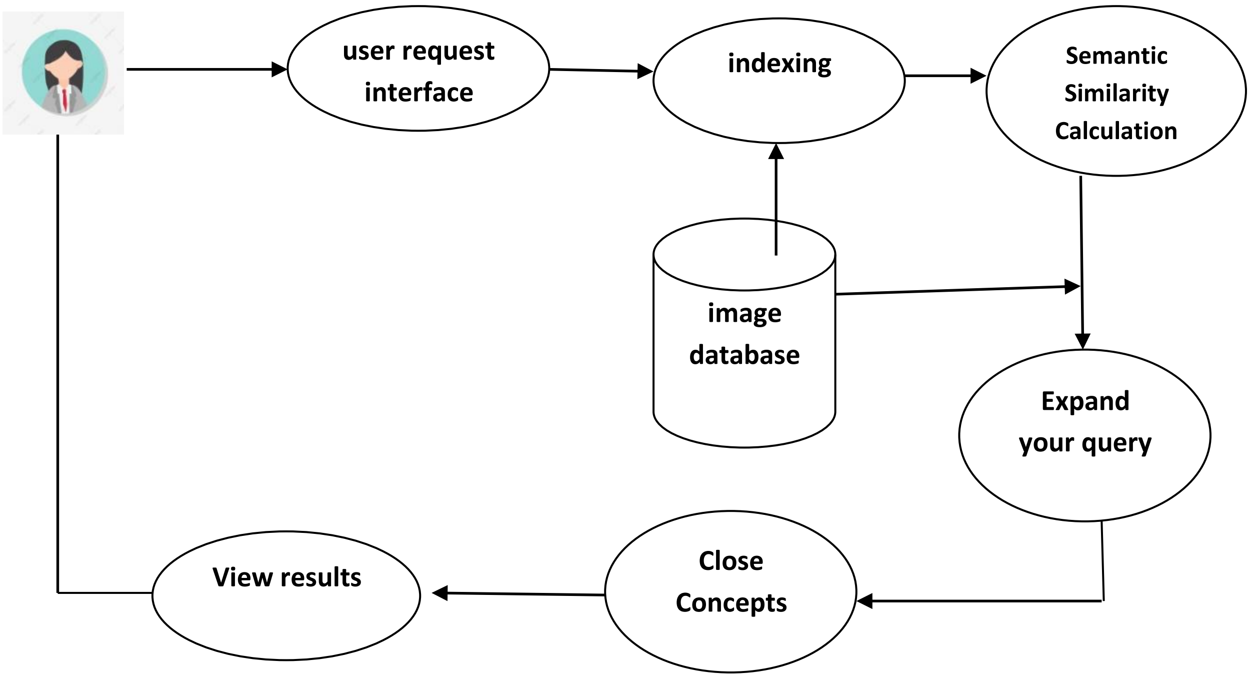


Figure 7: Query processing steps .

6- The algorithm:

Algorithm

1:Begin

2 :INPUT:Q = { Animals }

3: Compute the SSbetween Cn and all concepts in data base

$$sim_{res} = IC\ (LCS)$$

$$\text{sim}_{\boldsymbol{lin}} = \frac{\mathbf{2} \times \boldsymbol{IC}(\boldsymbol{LCS})}{\boldsymbol{IC}_{\boldsymbol{c1}} + \boldsymbol{IC}_{\boldsymbol{c2}}}$$

4: Query expansion: Wild animals (Lion , wolf ,tigr, panda, Bear…)

Domestic animals(Cat, sheep, Rabbit…)

5: Extract all images related toCn

6:OUTPUT: Result of images I_i(……..)

7:End

Example :

In this example, we provide a simple explanation for this type of research, as this research focuses on retrieval of images on semantic similarity, as it helps the user to find his shadow even if he does not know it by means of the similarity that brings him closer to the concept and the connotations that are from the same family and have the same characteristics.

Example 1 : Animals

The closest concepts:

Wild animals

Lion wolf tigr panda Bear





Domestic animals

Cat sheep Rabbit



Results :

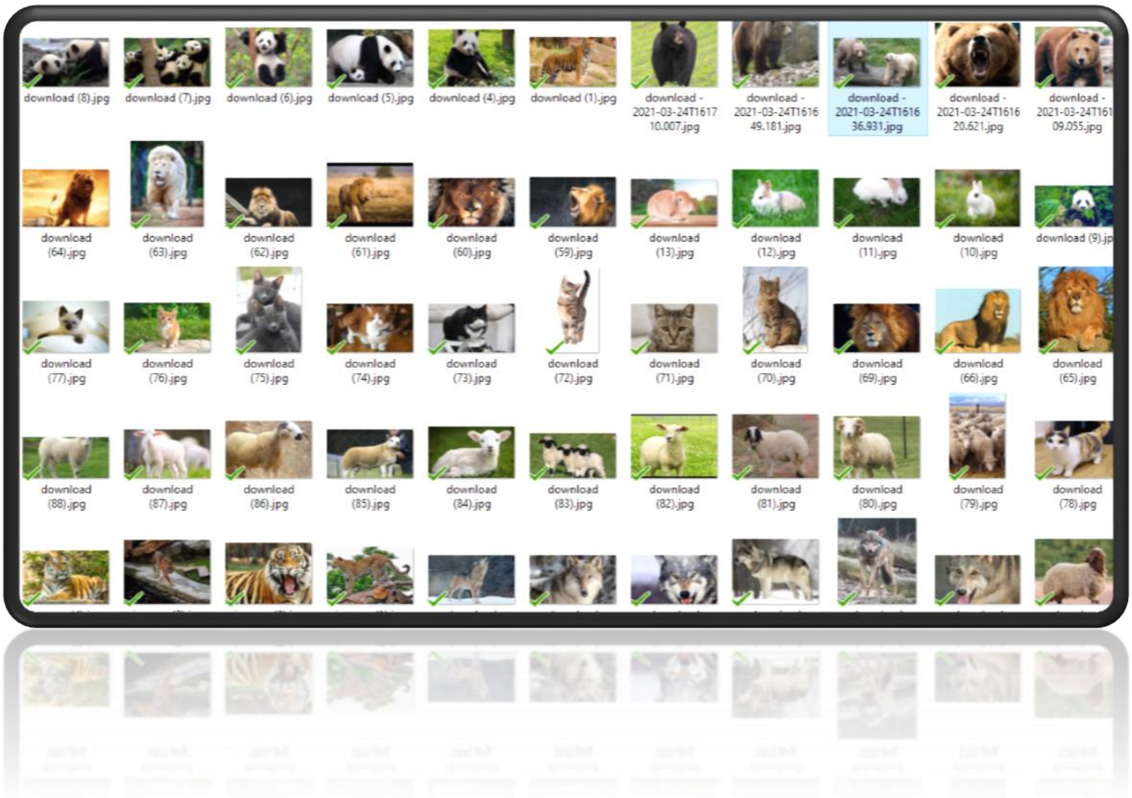


Figure 8 : All images are related to the concept**.**

7- Conclusion:

In this chapter, we talked about semantic approaches to expanding the query, as we first presented an overview of expanding the query, which is one of the solutions presented to bridge the gap of the original query and its purpose is to improve the retrieval rate, and secondly, we learned about the semantic similarity as it is based on the idea of distance between the elements and the similarity of their meaning or semantic content. It is used in specific applications such as information retrieval, recommendation systems, natural language processing, etc. We dealt with identifying a special case of information retrieval by semantic similarity in image retrieval. This type of research relies on semantic concepts as well as on the similarities between stored concepts.

The semantics behind each phrase were identified thanks to the utilization of semantic similarity to clarify and acquire the most likely target of the user's keyword.It is of great importance in expanding the search request. It helps to discover the actual meaning of each of the keywords, as it allows the user to access the appropriate information system in all transparency, taking into account the different characteristics that the accessed systems may show .

Chapter III:

Experimentation and Validation

1- Introduction:

After studying the different axes to understand our approach, we present in this chapter all the stages of design and implementation of our image search engine.

In this part we describe our search engine and explain our search model that works in the field of ontology to find more images that are relevant to user needs

Experimental setup

2- Visual Studio:

Microsoft Visual Studio is Microsoft's main integrated development environment. It enables graphical user interface programming and scripts as well as Windows Form, websites, web applications, and web services supported by Microsoft Windows, Windows Mobile, the .NET Framework, and Microsoft Silverlight .

Visual Studio contains a code editor that supports Intelligence technology and code rewriting, and also contains a translator that detects runtime errors and an interpreter that detects spelling errors in the code and also contains a model designer to build a graphical user interface, a web designer, classes designer (class (computer science)) and a grammar chart designer Data and designer for crystal reports .

Visual Studio supports many programming languages such as Microsoft Visual C ++, Microsoft Visual C #, Microsoft Visual Basic, JavaScript and also many markup languages such as HTML, XML, Extensible HTML, and XML .

.

2-1 Specifications:

2-1-1- Code Editor:

Visual Studio contains a code editor that supports formula instruction and automatic completion and also uses Intelligence technology to help the programmer write variables, functions, methods and cycles quickly, and the editor supports it in writing all programming and coding languages that Visual Studio contains .

The Visual Studio code editor also supports the ability to bookmark in the code to help with fast browsing, it also supports the ability to collapsing and expanding code collections, search and replace in the code, and it also supports code snippets, which are ready-madetemplates of code that can be inserted and changed into the projects in progress. Also the ability to rewrite the code Old motto**,**The code editor also puts red lines underneath the misspellings and green lines underneath the alerts.

2-1-2- Bug Tracker:

Visual Studio contains an error tracker that is supported by all supported languages. It detects runtime errors and spelling errors and allows placing breakpoints on lines of code that the program stops working when it reaches this line .

In Visual Studio there is also an immediate window that allows functions to be tested while writing them.

3-Design of our Ontology:

We have chosen 4 fields for our ontology field (nature, means of transportation, library and fashion). In the first step is to choose a group of pictures that belong to our field, then we use the ontology to comment on the pictures, the goal is to improve the quality of the search when the user returns

This work will allow modeling of the relationship between a subclass and its parent class.

Concept	mother concept	
Desert (c1)	nature	inanimate being
Forest (c2)		
The sea (c3)		
frozen pole (c4)		
sea animals	nature	Be alive
sand (c5)	Desert	
Palm (c6)	Desert	Plant them deser
dates (c7)		
Fenk (c8)		Animals Desert
Desert gazelle (c9)		
Camel (c10)		
Flowers (c11)	forest	Les plante
Tree (c12)		
River (c13)	The sea	
Dolphin (c14)	The sea	sea animal
Fish (c15)		
Shark (c16)		
Polar bear(c17)	frozen pole	Pole animals
Penguin (c18)		
Panda (c19)	Animals	Wild animals
Bear (c20)		
Lion (c21)		
Tiger (c22)		
Gazell (c23)		
Loup (c 24)		
Rabbit (c25)	Animals	Domestic animals
Chat (c26)		
Sheep (c27)		

concept	mother concept
A library (c28)	Un bibliothéque
pencil (c29)	
a chair (c30)	
table (c31)	
Book (c32)	
Desk (c33)	
University (c34)	
shelves (c35)	

concept	mother concept
car (c36)	road transport
bicycle(c37)	
Motorcycle (c38)	
Train(c39)	
a ship (c40)	Maritime transportation
Boat (c41)	
Plane (c42)	Air transportation
Helicopter (c43)	

concept	mother concept
Sewing machine (c44)	designs
A pair of scissors (c45)	
Sewing thread (c46)	
Sewing material (c47)	
clothes designing (c48)	

The following diagram graphically represents our Ontology (classes and class hierarchy of the ontology)

Ontology

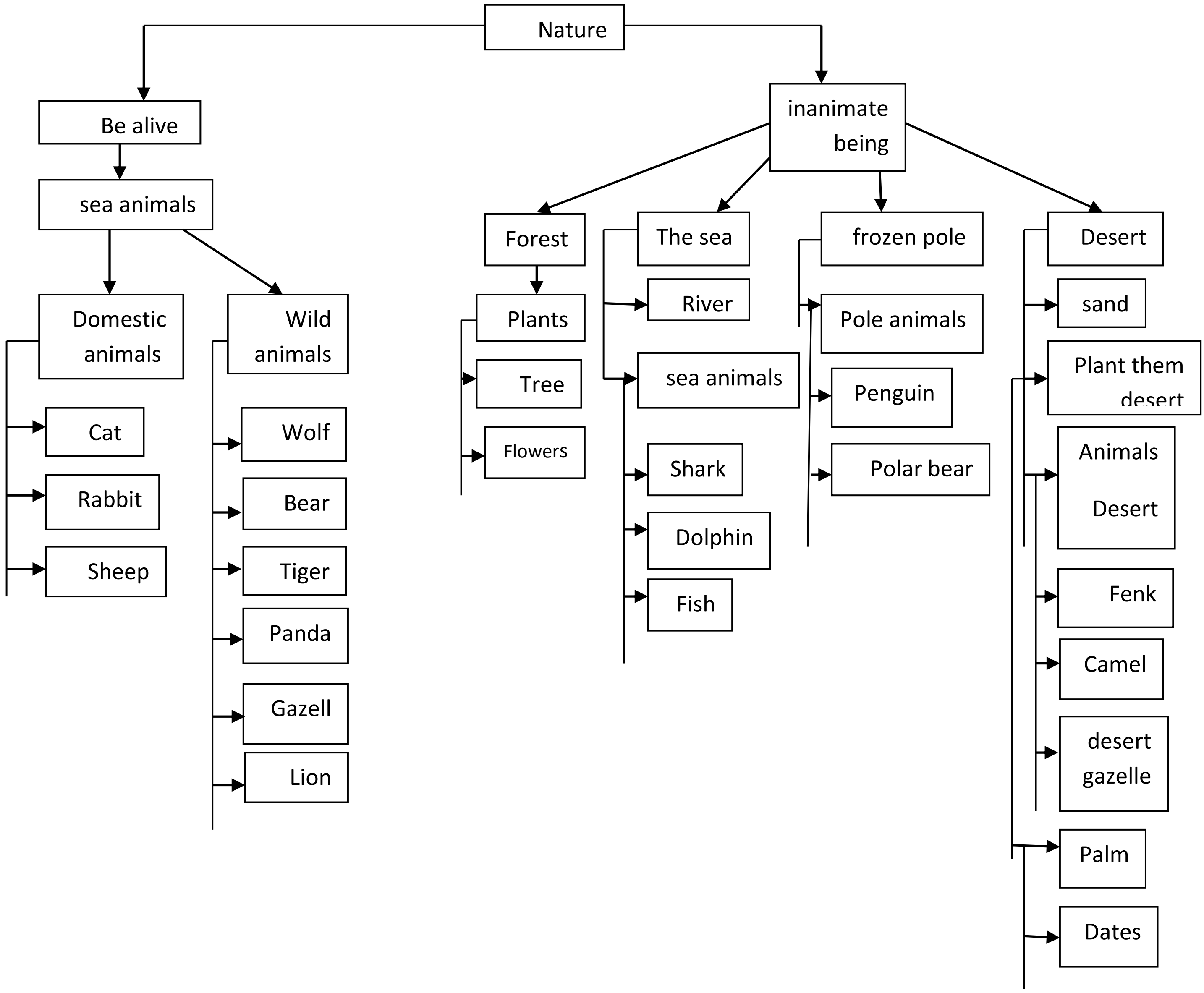


Figure 9: our ontology1.

Ontology

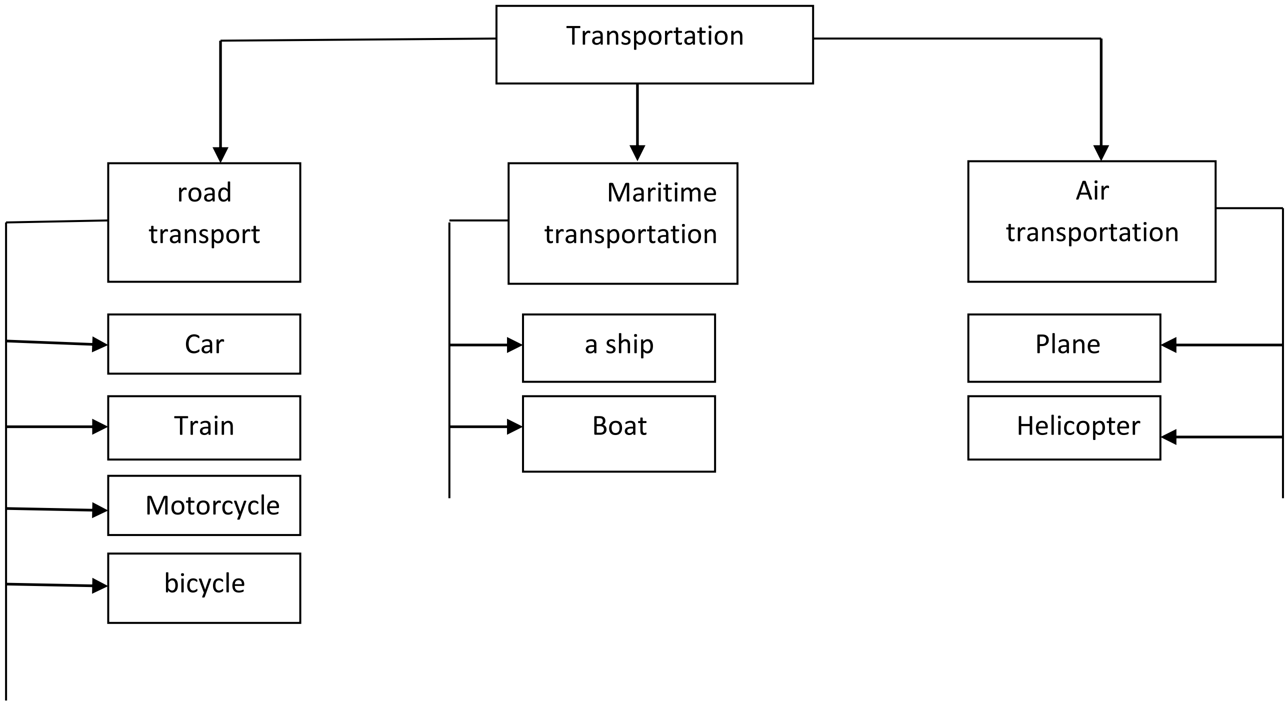


Figure 10: our ontology2.

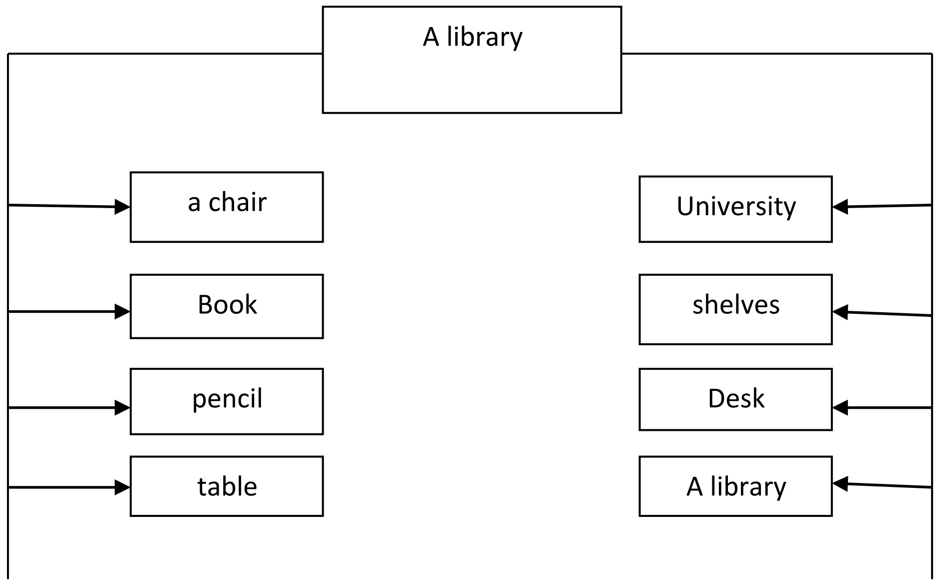


Figure 8: our ontology3.

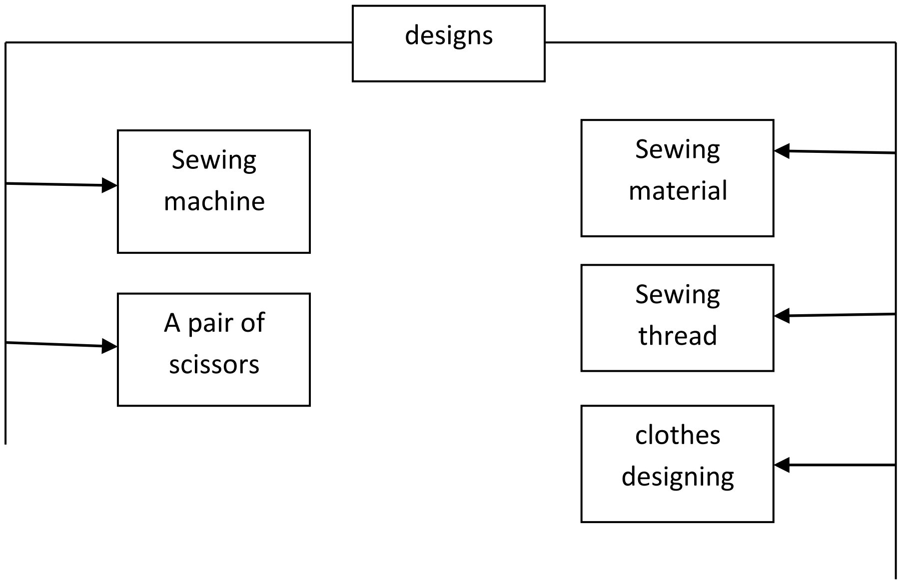


Figure 9: our ontology4.

4- General Matrix:

1- We have 48 concepts

2- In calculating the similarity between concepts, we used the Resnik measure and the Lin . measure that we talked about in Chapter 2:

Measure Resnik $sim_{res} = IC\ (LCS)$

$$sim_{res} = -\log P_{(c)} \qquad \rightarrow\ P = \frac{F}{N}$$

F = concept frequency

N = The total number of the concept

Measure Lin $\mathrm{sim}_{\boldsymbol{lin}} = \frac{\mathbf{2} \times \boldsymbol{IC}(\boldsymbol{LCS})}{\boldsymbol{IC}_{\boldsymbol{c1}} + \boldsymbol{IC}_{\boldsymbol{c2}}}$

Nature: 27 concept

A libray : 8 concept

Transportation : 8 concept

Disigne: 5 concept

	C1	C2	C3	C4	C5	C6	C7	C8	C9	0C1	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
C1		0.49	0.52	0.51	0.43	0.52	0.51	0.43	0.43	0.43	0.52	0.51	0.49	0	0	0	0	0	0	0	0.51	0.54	0	0.54	0.51	0.49	0.52
C2	0.49		0.60	0.58	0.49	0.60	0.58	0.49	0	0	0.60	0.58	0.55	0	0	0	0	0	0.58	0.52	0.58	0.62	0.55	0.62	0.58	0.55	0.60
C3	0.52	0.60		0.63	0.52	0	0	0	0	0	0	0	0.60	0.63	0.68	0.63	0	0	0	0	0	0	0	0	0	0	0
C4	0.51	0.58	0.63		0	0	0	0	0	0	0	0	0	0	0.65	0	0.63	0.63	0	0	0	0	0	0	0	0	0
C5	0.43	0.49	0.52	0		0.52	0.51	0.43	0.43	0.43	0.52	0.51	0.49	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C6	0.52	0.60	0	0	0.52		0.63	0.52	0.52	0.52	0.66	0.63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C7	0.51	0.58	0	0	0.51	0.63		0.51	0.51	0.51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C8	0.43	0.49	0	0	0.43	0.52	0.51		0.43	0.43	0	0	0	0.51	0.54	0.51	0.52	0.52	0.51	0.46	0.51	0.54	0.49	0.54	0.51	0.49	0.52
C9	0.43	0	0	0	0.43	0.52	0.51	0.43		0.43	0	0	0	0.51	0.54	0.51	0.52	0.52	0.51	0.46	0.51	0.54	0.49	0.54	0.51	0.49	0.52
C10	0.43	0	0	0	0.43	0.52	0.51	0.43	0.43		0	0	0	0.51	0.54	0.51	0.52	0.52	0.51	0.46	0.51	0.54	0.49	0.54	0.51	0.49	0.52
C11	0.52	0.60	0	0	0.52	0.66	0	0	0	0		0.63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C12	0.51	0.58	0	0	0.51	0.63	0	0	0	0	0.63		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C13	0.49	0.55	0.60	0	0.49	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0
C14	0	0	0.63	0	0	0	0	0.51	0.51	0.51	0	0	0		0.65	0.61	0.63	0.63	0.61	0.54	0.61	0.65	0.58	0.65	0.63	0.58	0.63
C15	0	0	0.68	0.65	0	0	0	0.54	0.54	0.54	0	0	0	0.65		0.65	0.68	0.68	0.65	0.65	0.71	0.68	0.71	0.65	0.68	0.65	0.68
C16	0	0	0.63	0	0	0	0	0.51	0.51	0.51	0	0	0	0.61	0.65		0.63	0.63	0.61	0.54	0.61	0.65	0.58	0.65	0.61	0.58	0.63
C17	0	0	0		0	0	0	0.52	0.52	0.52	0	0	0	0.63	0.68	0.63		0.66	0.63	0.57	0.63	0.68	0.60	0.68	0.63	0.60	0.66
C18	0	0	0	0	0	0	0	0.52	0.52	0.52	0	0	0	0.63	0.68	0.63	0.66		0.63	0.57	0.63	0.68	0.60	0.68	0.63	0.60	0.66
C19	0	0.58	0	0	0	0	0	0.51	0.51	0.51	0	0	0	0.61	0.65	0.61	0.63	0.63		0.54	0.61	0.65	0.58	0.65	0.61	0.58	0.63
C20	0	0.52	0	0	0	0	0	0.46	0.46	0.46	0	0	0	0.54	0.65	0.54	0.57	0.57	0.54		0.54	0.58	0.52	0.58	0.54	0.52	0.57
C21	0.51	0.58	0	0	0	0	0	0.51	0.51	0.51	0	0	0	0.58	0.71	0.61	0.63	0.63	0.61	0.54		0.65	0.58	0.65	0.61	0.58	0.63
C22	0.54	0.62	0	0	0	0	0	0.54	0.54	0.54	0	0	0	0.65	0.68	0.65	0.68	0.68	0.65	0.58	0.65		0.62	0.71	0.65	0.62	0.68
C23	0	0.55	0	0	0	0	0	0.49	0.49	0.49	0	0	0	0.58	0.71	0.58	0.60	0.60	0.58	0.52	0.58	0.62		0.62	0.58	0.55	0.60
C24	0.54	0.62	0	0	0	0	0	0.54	0.54	0.54	0	0	0	0.65	0.65	0.65	0.68	0.68	0.65	0.58	0.65	0.71	0.62		0.65	0.62	0.68
C25	0.51	0.58	0	0	0	0	0	0.51	0.51	0.51	0	0	0	0.63	0.68	0.61	0.63	0.63	0.61	0.54	0.61	0.65	0.58	0.65		0.62	0.68
C26	0.49	0.55	0	0	0	0	0	0.49	0.49	0.49	0	0	0	0.58	0.65	0.58	0.60	0.60	0.58	0.52	0.58	0.62	0.55	0.62	0.62		0.60
C27	0.52	0.60	0	0	0	0	0	0.52	0.52	0.52	0	0	0	0.63	0.68	0.63	0.66	0.66	0.63	0.57	0.63	0.68	0.60	0.68	0.68	0.60	

Figure 10: Concept Matrix .

	C28	C29	C30	C31	C32	C33	C34	C35	C36	C37	C38	C39	C40	C41	C42	C43	C44	C45	C46	C47	C48
C28		0.72	0.68	0.68	0.84	0.78	0.81	0.72	0	0	0	0	0	0	0	0	0	0	0	0	0
C29	0.72		0.68	0.68	0.84	0.78	0.81	0.72	0	0	0	0	0	0	0	0	0	0	0	0	0
C30	0.68	0.68		0.64	0.78	0.73	0.76	0.68	0	0	0	0	0	0	0	0	0	0	0	0	0
C31	0.68	0.68	0.64		0.78	0.73	0.76	0.68	0	0	0	0	0	0	0	0	0	0	0	0	0
C32	0.84	0.84	0.78	0.78		0.92	0.96	0.84	0	0	0	0	0	0	0	0	0	0	0	0	0
C33	0.78	0.78	0.73	0.73	0.92		0.89	0.78	0	0	0	0	0	0	0	0	0	0	0	0	0
C34	0.81	0.81	0.76	0.76	0.56	0.89		0.81	0	0	0	0	0	0	0	0	0	0	0	0	0
C35	0.72	0.72	0.68	0.68	0.84	0.78	0.81		0	0	0	0	0	0	0	0	0	0	0	0	0
C36	0	0	0	0	0	0	0	0		0.71	0.82	0.79	0.82	0.75	0.75	0.82	0	0	0	0	0
C37	0	0	0	0	0	0	0	0	0.71		0.73	0.71	0.73	0.68	0.68	0.73	0	0	0	0	0
C38	0	0	0	0	0	0	0	0	0.82	0.73		0.82	0.85	0.78	0.78	0.85	0	0	0	0	0
C39	0	0	0	0	0	0	0	0	0.79	0.71	0.82		0.82	0.75	0.75	0.82	0	0	0	0	0
C40	0	0	0	0	0	0	0	0	0.82	0.73	0.85	0.82		0.78	0.78	0.85	0	0	0	0	0
C41	0	0	0	0	0	0	0	0	0.75	0.68	0.78	0.75	0.78		0.72	0.78	0	0	0	0	0
C42	0	0	0	0	0	0	0	0	0.75	0.75	0.78	0.75	0.78	0.72		0.78	0	0	0	0	0
C43	0	0	0	0	0	0	0	0	0.82	0.73	0.85	0.82	0.85	0.78	0.78		0	0	0	0	0
C44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0.41	0.41	0.38	0.44
C45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.41		0.41	0.38	0.44
C46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.41	0.41		0.38	0.44

Comment

Picture (1) represents the similarity measurement matrix between our ontology concepts. The results were recorded where the resulting similarity values range from 0 (no relationship) to 1 (identical) and the result varies with the rest of the concepts according to the degree of similarity

5- General architecture of our search engine:

Our search engine works like any normal search engine where we follow these steps:

5-1- The indexing step:

We must have our own database where in the first step of indexing we extract the properties of the images (write annotation) and store them in a special structure called the index.

We chose in the annotation to do it manually, as we annotated each image in our group and commented on it according to the concepts of the ontology .

these annotations are stored in two formats in two separate files with the following structure:

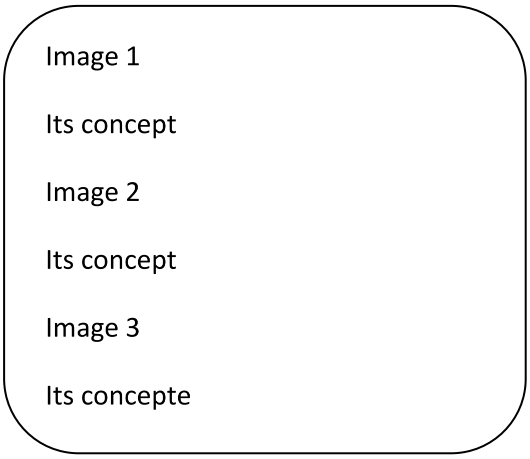


Figure 11: The image annotation file structure.

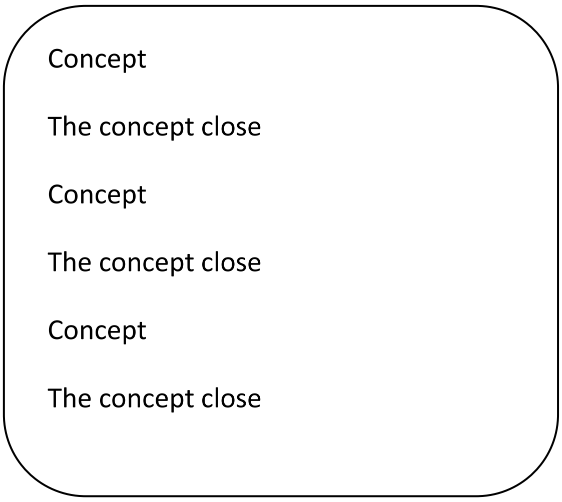


Figure 12: The image concept close file structure.

5-2- The Research Stage:

1- First, we formulate the request where the user chooses one of the semantic concepts where these concepts belong to different levels .

2- Second, the search engine displays the search models with the indexing query and then studies the similarity with our indexing base and displays the similarity indicators from the image base of our search engine.

3- Finally, display the search results, where the search engine displays the image results to the user in descending order from most to least similar to the query.

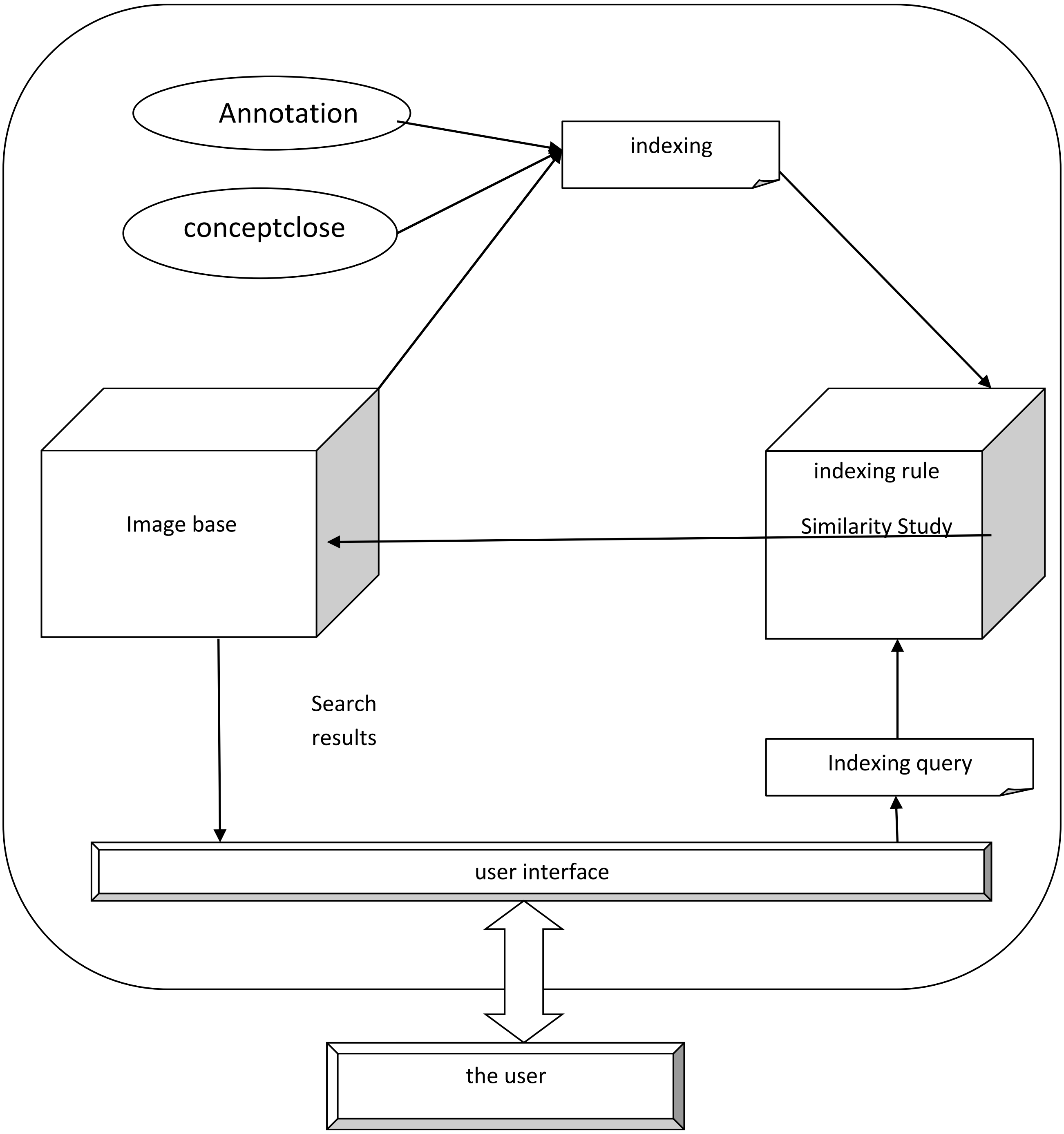


Figure 13: Engine architecture of our search .

6- Examples of our Search Engine:

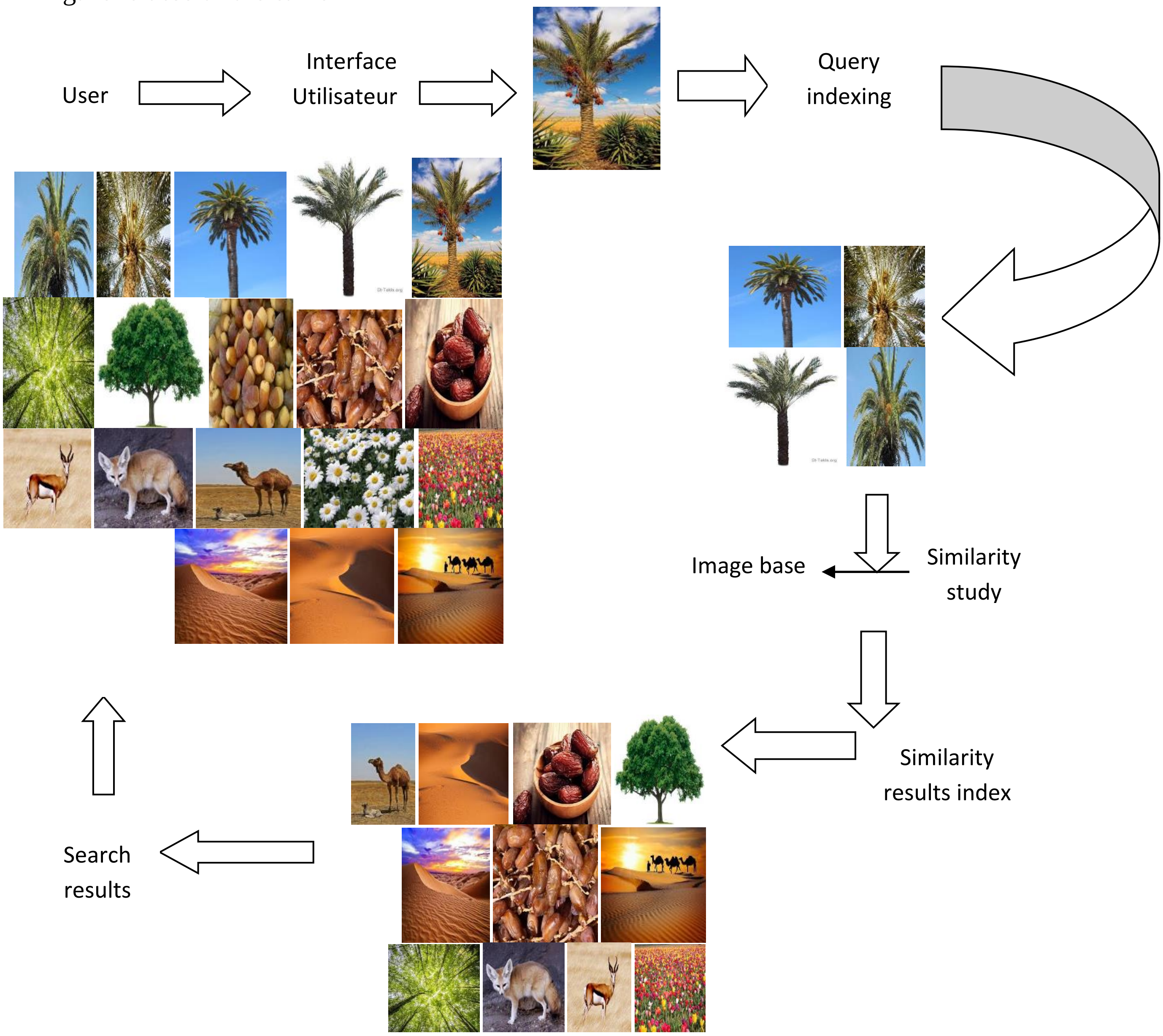
In our search engine, the user chooses one or more semantic concepts .

In the search phase a query is represented which is an analysis of user input and the search is expanded to match more data from our database .

Example 1 :

In this example, the user chooses a concept "**Palm**"by applying the Query Expansion QE, we find that

The concepts closest to the "**Palm**" are (the date - the sand - desert - tree - the flower - fennec - gazelle desert - the camel



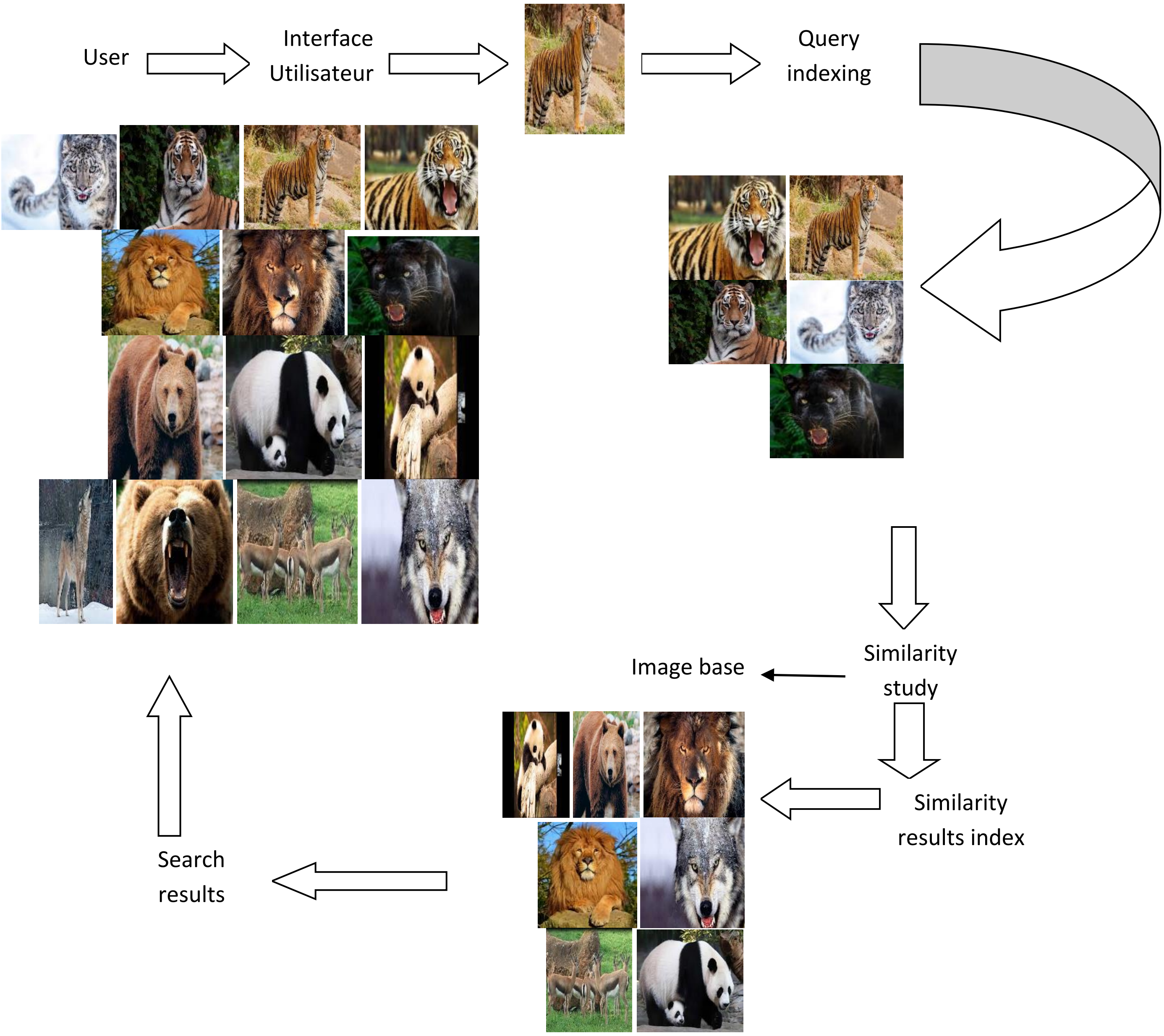
Semantic similarity:

	the date	the sand	the flower	tree	Fenk	gazelle desert	The camel	Desert
palm	**0.63**	**0.52**	**0.66**	**0.63**	**0.52**	**0.52**	**0.52**	**0.52**

Example 2 :

In this example, the user chooses a concept "**Tiger**",by applying the Query Expansion QE, we find that:

The concepts closest to the "**Tiger**"are(wild animal: lion - bear - gazelle - panda - wolf)



Semantic similarity:

	lion	Bear	panda	gazelle	wolf
Tiger	0.65	0.58	0.65	0.62	0.68

Experimental Results

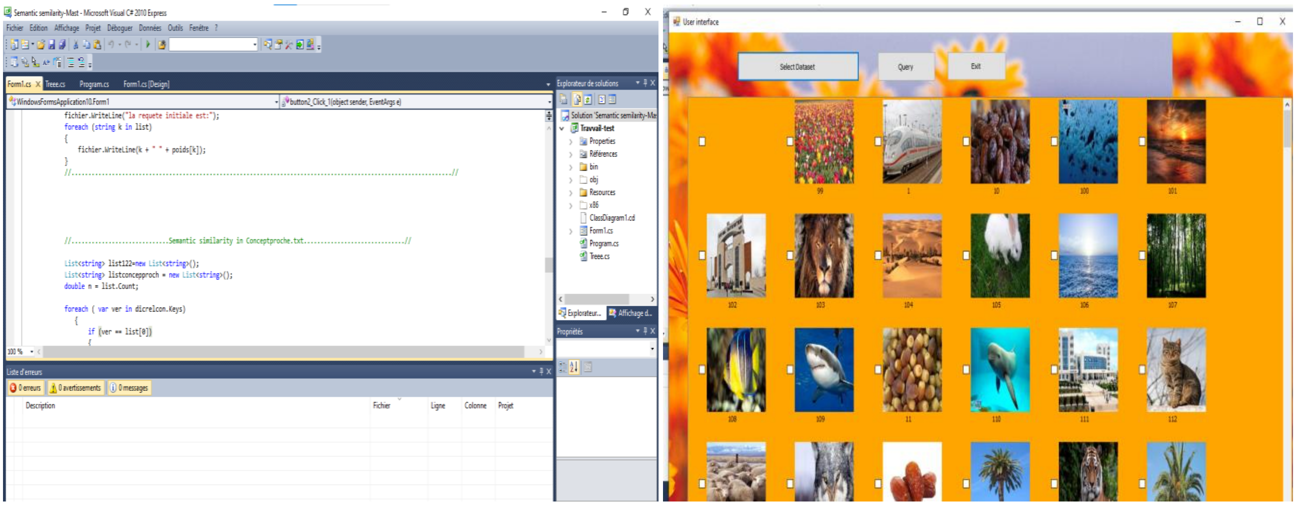


Figure 14: The main window of our application.

7- Experiment:

We implemented our search engine with C # language in the Visual studio environment

In our search engine experiment, we selected 10 people and asked each person to choose a concept from our image base. Once the request was submitted, our search engine searched for relevant images and then displayed them by calculating the similarity score.

The results of our experiment are in this table

Query	Relevant images retrieved in our algorithm	Relevant images retrieved in the conventional algorithm
	21	5
	23	5
	17	6
	21	6
	39	6
	39	8
	30	5
	19	4
	39	7
	23	7

Table 2: Comparison the results between our algorithm and the conventional algorithm..

8- Discuss the Results:

After conducting this experiment, we notice a big difference in the results obtained through Table (2). We find that our search engine gives us more results than the traditional search engine. Why?

Because our search engine depends in its search on semantic similarity and thus expands our search circle until we get the most possible and more accurate results, as it presents these results according to the percentage of similarity to the user's request .

On the grounds that the search engines agree between the user's query and the data contained in the system and retrieve the results corresponding to the query. We find that traditional search engines match the terms and are similar between the query and the text. While our search engine takes into account the meanings and semantics of terms and not only matches, so the results are more relevant to the user's query .

9- Conclusion:

In this chapter, we implement our algorithm. In the first part, we presented our ontology, then we explained how to calculate the general matrix of concepts(by using semantic similarity)after that we explained how our search engine works .

In the second part, we realized that our image search system exploits semantic metadata as our system allows us to provide images that are relevant to user intention. we compared our results with the previous work, We found that our algorithm clearly outperformed.

CONCLUSION

GENERAL

Conclusion General:

Image search is a branch of computer vision that entails navigating, searching, and retrieving images from a huge collection of digital images. It's a type of data search used to locate photos. As a result, picture indexing and search technologies that are specifically designed to store, organize, and find images quickly, efficiently, and effectively are increasingly becoming a must.

Several search engines, such as Google [69] and Lycos [70] offer image searches based on text they are based on the fundamental principle that in a web page, there must be content. There is a strong correlation between the text and the images. The main issue with these keyword searches is that the results can be completely irrelevant.

To characterize images, the first image search systems used keywords associated with the images. Using text-based methods to find photos containing the keywords is no longer necessary because to this association of keywords.

With the revolution in the world of technology, the size of image bases has become very large (making digital cameras, scanners, webcams and mobile phones with high computing and storage capabilities). A new field of research so far they are working on developing and developing, and it is known as image search engines. These engines can be classified into two categories.

- The first category (Content-Based Image Retrieval: CBIR) makes use of the visual content of images: in this mode, the user is typically asked to select image examples that are similar to what they are looking for. The search is carried out by comparing the characteristics of the lower levels of the requête to those of the photographs in the collection.

- The second category makes use of the semantic associations between images. This mode allows the user to construct their request using a textual query. The search is carried out by comparing the terms in the requête with the textual annotations that represent the collection's photos.

In our work, we talked about how to retrieve images with semantic similarity, meaning that our research was focused on semantics it means the study of words in terms of semantic classifications that have accuracy and depth in meaning and depend on some contexts.

Advances in 'lexical semantics' have furthered the development of 'semantic field theory' and 'semantic network' or 'semantic circuits of networks', those strategies that organize words according to interrelated semantic meanings. It is worth noting that these words may be collected together (related to each other) according to certain criteria, for example: the word animals can be collected in relation to natural properties and so on by analogy.

Semantic search, or what is known as the semantic web, or the web with meaning, and this is intended to rely on software that can define what is meant by the data provided by the World Wide Web (the Web) using what is known as ontology concept maps.Semantic search is one of the branches of artificial intelligence, and it is a revolution in the world of the web, as it allows the browser or proxy programs to search for information, and thus the process of processing information is based on computers instead of humans.Semantic search is not a separate network in itself but an extension of existing research, where information has clear meanings while better enabling computers and individuals to work collaboratively.

In the framework of our work the goal was to implement an ontology-based semantic search engineTo take advantage of the semantic richness it offers.

Our work revolves around the following main stages:

●The first chapter provided an overview of how to retrieve images, where we studied the different classic search models that can be used in the field of image search.

●The second chapter focused on two elements whose purpose is to improve the retrieval rate. Expansion of the query and semantic similarity have the same goal, which is to obtain more accurate information and retrieve it in a reasonable time and reduce the error rate during the retrieval of information. We also learned about the laws of measuring semantic similarity that are widely used in many among the functions are areas of computer science, among which we can cite the automatic processing of bioinformatics. Information retrieval ..... It allows to determine the similarity between terms and concepts that have no synchronization.

●In the third chapter, we have built ontology in 4 different fields, which are directed terms from ontology in these areas that are used in the annotation stage and in the research stage. Finally, we implemented a semantic search engine based on ontology, which uses the vector model after adapting it to our needs.

References

References:

[1] R. Priyatharshini , S. Chitrakala,"Association based Image retrieval: A survey,"Springer-Verlag Berlin Heidelberge, pp. 17-26, 2013.

[2] Vaishali D. Dhale , A. R. Mahajan, Uma Thakur, "A Survey of Feature Extraction Methods for Image Retrieval," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012.

[3] B E Prasad; A Gupta; H-M Toong; S.E. Madnick (February 1987). "A microcomputer-based image database management system" (PDF). IEEE Transactions on Industrial Electronics. IE-34 (1): 83–8.

[4]. Banireddy Prasaad, Amar Gupta, Hoo-min Toong A Gupta, H-M Toong, S. E. Madnick, "A microcomputer-bassed image database management system", IEEE transactions on Electronics. IE-34, pp. 83-84, February 1987.

[5] Y. Rui, T.S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," Journal of visual communication and image representation, vol. 10, 1999, pp. 39–62.

[6] Styrman A., "Ontology-Based Image Annotation and Retrieval", M.S. Thesis, University of Helsinki, April 2005.

[7] Rajeev Agrawal , wiliam Groska, Farshad Fotouhi, "image Clustering Using Multimodal keywords", SAMT, pp. 113-123, 2006

[8] Kherfi M.L., Ziou D., Bernardi A., "Image Retrieval from the World Wide Web: Issues, Techniques and Systems", ACM Computing Surveys, Vol.36, Issue 1, March 20004, pages 35-67

[9] Styrman A., "Ontology-Based Image Annotation and Retrieval", M.S. Thesis, University of Helsinki, April 2005.

[10] C. Faloutsos et al, 1994, "Efficient and effective querying by image content," Journal of intelligent information systems, Vol.3, pp.231-262.

[11] HuiHui Wang, DzulkifliMohamad, N.A. Ismail "Approaches, Challenges and Future Direction of Image Retrieval," Journal of Computing, Volume 2, Issue 6, June 2010, ISSN 2151-9617.

[12] Étude comparative des fonctionnalités des moteurs de recherche d'images sur Internet, Boudry Christophe et Agostini Clémence [archive], Documentaliste-Sciences de l'Information, 2004/2 (Vol. 41), p. 96-105. DOI 10.3917/docsi:.412.0096. (Consultation le 28 juin 2019).

[13] P.Jayaprabha, Rm.Somasundaram," Content Based Image Retrieval Methods Using Graphical Image Retrieval Algorithm (GIRA)," International Journal of Information and Communication Technology Research, Volume 2 No. 1, January 2012

[14] Feng Jing , Bo Zhang, Fuzong Lin , Wei-Ying Ma, Hong-Jiang Zhang "A Novel Region-Based Image Retrieval Method Using Relevance Feedback," This work was performed at Microsoft Research China 2005

[15] Derek Hoiem , Rahul Sukthankar , Henry Schneiderman , Larry Huston" Object-based image retrieval using the statistical structure of images," Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2004

[16]Masashi, "Image Retriveal: research and use in the information Explosion," Progress in informatics, No. -6, pp. 3-14, (2009).

[17] NidhiSinghai, Prof. Shishir K. Shandilya," A Survey On: Content Based Image Retrieval Systems," International Journal of Computer Applications (0975 – 8887) Volume 4 – No.2, July 2010

[18] GauravJaswal (student), AmitKaul," Content Based Image Retrieval – A Literature Review," National Conference on Computing, Communication and Control (CCC-09), 2009.

[19] MicheleSaad," Image Retrieval Literature Survey," EE 381K: Multidimensional Digital Signal Processing March 18, 2008.

[20]AmandeepKhokher, Dr. Rajneesh Talwar, "Content Based Image Retrieval: State-of-the-Art and Challenges," International Journal of Advanced Engineering Sciences and Technologies, Vol No. 9, Issue No. 2, 207-211

[21]YongRui and Thomas S. Huang, “Image Retrieval: Current Techniques, Promising Directions, and Open Issues,” Journal of Visual Communication and Image Representation 10, 39–62 1999

[22] AmandeepKhokher, Dr. Rajneesh Talwar, “Content Based Image Retrieval: State-of-the-Art and Challenges,” International Journal of Advanced Engineering Sciences and Technologies, Vol No. 9, Issue No. 2, 207-211

[23] Xin Jin, Jiebo Luo, Feb 2013, “Reinforced Similarity Integration in Image-Rich Information Networks” IEEE Trans. On Knowledge and Data Engineering, Vol.25, No.2, pp. 448-460.

[24] Kherfi M.L., Ziou D., Bernardi A., “Image Retrieval from the World Wide Web: Issues, Techniques and Systems”, ACM Computing Surveys, Vol.36, Issue 1, March 20004, pages 35-67.

[25] Rasiwasia, NikhilMoreno, Pedro J.Vasconcelos , Nuno, . Query by semantic example. University of California San Diego, Book p925 , 2007-08-01 ISSN 1520-9210

[26] R. Duda, P. Hart, and D. Stork, Pattern Classification. New York: Wiley, 2001.

[27] A. Dempster, N. Laird, and D. Rubin, “Maximum-likelihood from incomplete data via the EM algorithm,” J. R. Statist. Soc., vol. B-39, 1977.

[28] Semantic Concept-Based Query Expansion and Re-ranking for Multimedia Retrieval∗ A Comparative Review and New Approaches

[29] E. M. Vorhees. Query expansion using lexical-semantic relations. In 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 61–69, Dublin, Ireland, August 1994.

[30] Y. Qiu and H. Frei. Concept based query expansion. In Proc. 16th Annual Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval, pages 160–169, 1993

[31] K. S. Jones. Automatic Keyword Classification for Information Retrieval. Butterworths, 1971.

[32] 0] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(6):391–407, 1990.

[33] J. J. Rocchio. Relevance feedback in information retrieval, chapter 14. Relevance Feedback in Information Retrieval, pages 313–323. Prentice-Hall Inc., Englewood Cliffs, NJ, 1971.

[34] J. Xu and W. B. Croft. Query expansion using local and global document analysis. In 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 4–11, New York, NY, 18–22 August 1996.

[35] USING A QUERY EXPANSION TECHNIQUE TO IMPROVE DOCUMENT RETRIEVAL (Abdelmgeid Amin Aly) Article in International journal (Toronto, Ont.) · January 2008 page 3

[36] Survey of Automatic Query Expansion for Arabic Text Retrieval (JISTaP http://www.jistap.org Journal of Information Science Theory and Practice eISSN : 2287-4577 pISSN : 2287-9099) page 69.

[37] (Sharma, Pamula, & Chauhan, 2019) Sharma, D., Pamula, R., & Chauhan, D. S. (2019). A hybrid evolutionary algorithm based automatic query expansion for enhancing document retrieval system. Journal of Ambient Intelligence and Humanized Computing, 1-20.

[38] Han and Chen (2009) Han, L., & Chen, G. (2009). HQE: A hybrid method for query expansion. Expert Systems with Applications, 36(4), 7985- 7991.

[39] - Budanitsky, A. and Hirst, G. 2006. “Evaluating WordNet-based measures of semantic distance,” Comput. Linguistics, vol. 32, no. 1, pp. 13–47.

[40] Harispe S.; Ranwez S. Janaqi S.; Montmain J. (2015). "Semantic Similarity from Natural Language and Ontology Analysis". Synthesis Lectures on Human Language Technologies. 8:1: 1–254.

[41] (Tchechmedjev, 2012) Tchechmedjev A., tat de l’art : mesures de similarit smantique locales et algorithmes globaux pour la dsambigusation lexicale base de connaissances , Actes de la confrence conjointe JEP-TALN-RECITAL, p. pp : 295-308, 2012.

[42] Etude des mesures de similarit ´ e s ´ emantique ´ basees sur les arcs (Aly Ngone Ngom ´ 1 LANI (Laboratoire d'Analyse Numerique et d'Informatique) ´ Universite Gaston Berger ´ BP 234, Saint-Louis, Sen´ egal ´ alyngonengom@gmail.com)page 1/2/3/4/5/6 .

[43] Computing Semantic Similarity of Concepts in Knowledge Graphs (Ganggao Zhu and Carlos A. Iglesias/I EEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. X, XXXXX 2016) page 3 /4.

[44] K. W. Church and P. Hanks, "Word association norms, mutual information, and lexicography," Comput. Linguist., vol. 16, no. 1, pp. 22–29, Mar. 1990 .

[45]) R. Gligorov, W. ten Kate, Z. Aleksovski, and F. van Harmelen, "Using google distance to weight approximate ontology matches," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 767– 776 .

[46] T. K. Landauer and S. T. Dumais, "A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge," Psychological Rev., vol. 104, no. 2, 1997, Art. no. 211 .

[47]E. Gabrilovich and S. Markovitch, "Computing semantic relatedness using wikipedia-based explicit semantic analysis," in Proc. 20th Int. Joint Conf. Artif. Intell., 2007, vol. 7, pp. 1606–1611.

[48] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Advances in Neural Information Processing Systems 26, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds. Red Hook, NY, USA: Curran, 2013, pp. 3111– 3119

[49] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in Proc. Empiricial Methods Natural Language Process., 2014, vol. 12, pp. 1532–1543.

[50] P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics," J. Artif. Intell. Res., vol. 37, no. 1, pp. 141–188, 2010 .

[51] R. Rada, H. Mili, E. Bicknell, and M. Blettner, "Development and application of a metric on semantic nets," IEEE Trans. Syst. Man Cybernetics, vol. 19, no. 1, pp. 17–30, Jan./Feb. 1989

[52] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," in Proc. 21st Nat. Conf. Artif.Intell., vol. 6, 2006, pp. 775–780.

[53] (Rada et al., 1989) Rada R., Mili H., Bicknell E., Blettnerr M., Development and application of a metric on semantic nets , IEEE Transaction on systems, Man, and Cybernetics 19(1)p. pp : 17 - 30, 1989.

[54] C. Leacock and M. Chodorow, "Combining local context and WordNet similarity for word sense identification," WordNet: Electron. Lexical Database, vol. 49, no. 2, pp. 265–283, 1998.

[55] (Leacock et Chodorow, 1998) Leacock C., Chodorow M., Combining local context and WordNet sense similarity for word sense identification , In WordNet, An Electronic Lexical Database. The MIT Press, 1998.

[56] . Wu and M. Palmer, "Verbs semantics and lexical selection," in Proc. 32nd Annu.Meeting Assoc. Comput. Linguistics, 1994, pp. 133– 138

[57] (Wu et Palmer, 1994) Wu Z., Palmer M., Verbs semantics and lexical selection , in U. A. f. C. L. Stroudsburg, PA (ed.), In Proceedings of the 32nd annual meeting on ACL, volume 2 de ACL '94, p. pp : 133-138, 1994

[58] (Sussna, 1993) Sussna M., Word sense disambiguation for free-text indexing using a massive semantic network , in Proc. of the Second International Conference on Information and Knowledge Management, p. pp : 67-74, 1993.

[59] (Hirst et St-Onge, 1998) Hirst G., St-Onge D., Lexical chains as representations of context for the detection and correction of malapropisms , In Fellbaum, 1998 The MIT Pressp. pp : 305-332, 1998.

[60] (Stojanovic et al., 2001) Stojanovic N., Maedche A., Staab S., Stuber R., Sure Y., Seal : a framework for developing semantic portals , in Proc. of the int. conf. on Knowledge capture, p. pp : 155-162, 2001.

[61] **(**Zhong et al., 2002) Zhong J., Zhu H., Li J., Yu Y., Conceptual graph matching for semantic search , in Proceedings of the 10th International Conference on Conceptual Structures (ICCS'02) (London, UK), Springer-Verlag, p. pp : 92-106, 2002 .

[62] (Zargayouna (2004) Zargayouna H., Contexte et smantique pour une indexation de documents semi-structurs , LIMSI/CNRS-Universit Paris 11p. pp : 1-15, 2004.

[63] (Slimani et al., 2007) Slimani T., Yaghlane B. B., Mellouli K., Une extension de mesure de similarit entre les concepts d'une ontologie , 4th International Conference : Science of Electronic Technolgies of Information and Telecommunications, SETIT 2007 March 25 -29 2007 - Tunisia, p. pp : 1 - 10, 2007.

[64] (Resnik, 1995) Resnik P., Using Information Content to Evaluate Semantic Similarity in a Taxonomy , In Proceedings of 14th International Joint Conference on Artificial Intelligence, Montreal, 1995.

[65] Measuring the Semantic Similarity of Texts Courtney Corley and Rada Mihalcea Department of Computer Science University of North Texas {corley,rada}@cs.unt.edu p14

[66] (Lin, 1998) . Lin and E.H. Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In Proceedings of Human Language Technology Conference (HLTNAACL 2003), Edmonton, Canada, May .

[67] (Jiang and Conrath, 1997) Jiang and D. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In Proceedings of the International Conference on Research in Computational Linguistics, Taiwan.

[68] Content-based Image Retrieval (John Eakins Margaret Graham University of Northumbria at Newcastle page 25 -26 .

[69] Moteur de recherche web Google. http://www.google.fr

[70] Moteur de recherche web Lycos. http://www.lycos.f