



N° : .....

**Msila University**  
**FACULTY OF MATHEMATICS ET AND COMPUTER**  
**SCIENCE**

**Computer Science Department**

**Being a dissertation submitted for the degree of MASTER of  
Computer Science**

**Domain: Mathematics and Computer Science(AIS)**

**Branch: Computer Science**

**Specialty: Advanced Information Systems**

**By : BAGHDADI Ammar**

**Title:**

**Information Retrieval in Multimedia Documents**

**Presented publicly in: / /2013.**

.....

**Msila University**

**President**

.....

**Msila University**

**Rapporteur**

.....

**Msila University**

**Examiner**

.....

**Msila University**

**Examiner**

**Promotion : 2012 /2013**

## Contents

General Introduction .....	3
I. CHAPTER 1.....	4
Multimedia Retrieval and Digital Libraries.....	4
1.1. Introduction.....	4
1.2. Brief history.....	4
1.3. Motivation.....	5
1.3.1. Multimedia information .....	5
1.3.2. Multimedia libraries.....	5
1.4. Fundamental Concepts and Definitions.....	6
1.4.1. Document and Collection of Documents .....	6
1.4.2. Contextual IR is not Content-based IR.....	7
1.5. Views on Multimedia Documents.....	7
1.6. Methods of Information Retrieval.....	8
1.6.1. Free Browsing.....	9
1.6.2. Text-based information retrieval .....	9
1.6.3. Content-based information retrieval.....	10
1.6.3.1. Content-based image retrieval (CBIR) .....	11
1.6.3.2. Content-based audio retrieval (CBAR) .....	13
1.6.3.3. Content-based video retrieval (CBVR) .....	14
1.7. Conclusion.....	16
II. CHAPTER 2: .....	17
System Design and Specifications.....	17
2.1. Introduction.....	17
2.2. Metadata.....	17
2.2.1. Metadata for multimedia.....	18
2.2.1.1. Semantic metadata.....	18
2.2.1.2. Context metadata.....	19
2.2.1.3. Structural metadata.....	19
2.2.2. Multiple metadata documents.....	19

2.2.3.	Characteristics of text documents.....	19
2.2.4.	Characteristics of images.....	20
2.3.	Indexing Media DBs.....	22
2.4.	Features specification.....	23
2.4.1.	Image features specification.....	23
2.4.2.	Audio features specification.....	23
2.4.3.	Video features specification.....	23
2.5.	System Design.....	24
2.5.1.	Use Case Diagram.....	24
2.5.2.	Data Flow Diagram.....	25
2.5.3.	Block Diagram.....	26
2.5.4.	Deployment Diagram.....	27
2.5.5.	Conclusion.....	28
III.	CHAPTER 3.....	29
	Implementation and testing .....	29
3.1.	Introduction .....	29
3.2.	The environment of development.....	29
3.3.	Dominant color Problematic.....	30
3.4.	Proposed Dominant Color Name technique.....	31
3.4.1.	Color Names.....	32
3.4.2.	Converting RGB to hexadecimal.....	32
3.5.	Implementation of the proposed algorithm.....	33
3.6.	Metadata extraction.....	34
3.7.	The application .....	35
3.7.1.	Data acquisition.....	35
3.7.2.	Search.....	36
3.8.	Conclusion.....	39
	General Conclusion.....	40
	Bibliography .....	41

## **Introduction:**

Information is increasingly becoming ubiquitous and all pervasive, throughout multimedia repositories. With the popularity of multimedia technology documents, they have been a lot more versatile than a few years ago. However, although more information is available on the web, the efficient and effective retrieval and management of these multimedia documents are still very challenging research issues. When searching in such a vast collection of multimedia documents, users can easily get lost in its depths. Some users know what they are looking for and try to satisfy their needs by refining keywords. Those users may or may not find something of interest, but may easily miss other. Other users who know what they want can express their needs to a software called a *search engine*, which, ostensibly, helps them find the appropriate documents.

Search engines are still in their infancy. Although the exact nature of many of the algorithms they use for finding appropriate results is proprietary [G. Pringle] This dissertation tries to build a Multimedia Documents Search Engine software that helps users retrieving multimedia documents from multimedia digital libraries to fit user need and to give many choices of search set for refining requested results. So our work is placed in the context of multimedia information retrieval (MIR), by means, searching for documents themselves, searching for metadata which describe documents, or searching within databases.

This dissertation is structured in three chapters, the first chapter dedicated for viewing principal concepts and approaches on Multimedia Retrieval and Digital Libraries and then citation of the different methods to retrieve requested documents, in the second chapter Multimedia documents search Engine Software will be designed including all important diagrams, plus, all the features to deal with will be specified and limited to precise the requested purpose. The last chapter will be the demonstration of the proposed Multimedia documents search Engine Software. And reviewing the proposed dominant color algorithm.

## **General conclusion**

In this dissertation, we have examined the use of Classify algorithm for content-based document retrieval, using both keywords and the color of the image document. As well as querying into metadata on audio documents and video documents although the search process doesn't support all formats and all multimedia document types. However, this Multimedia search engine is not efficient enough for most types of applications. Furthermore, in many applications, only some of the concepts described here are necessary. On the other hand, many database management issues are not addressed in our system. so in the future work, as I said above, I'll be interested in Aforge.net Framework library that opened great chance for developing successful search Engine with great effectiveness and efficiency.

## Bibliography

- [1] B. V. Patel and B. B. Meshram. (October 2012). CONTENT BASED VIDEO RETRIEVAL. *The International Journal of Multimedia & Its Applications (IJMA)*.
- [2] Dornfest,R. and Brickley,D. (2001, last viewed Sept.2006). *The Power of Metadata*. O'Reilly Network.
- [3] Dunlop, M. D. (1991). *Multimedia Information Retrieval*. Glasgow.
- [4] G. Pringle, L. Allison, and D. L. Dowe, "What is a tall poppy among web pages?," in Proc. 7th IWWWC, Brisbane, Australia, 1998.
- [5] Hillmann, D. (2005). *Using Dublin Core. Dublin Core Metadata Initiative*.
- [6] Liu, G. (. (2008). *Content-Based Information Retrieval and Digital Libraries*. Chicago: Information Technology and Libraries.
- [7] LANCASTER, F.W. (1968) *Information Retrieval Systems: Characteristics, Testing and Evaluation*, Wiley, New York.
- [8] Martínez, R. M. (novembre 2009). *An introduction to content-based Information Retrieval by Normalized Compression Distance*. AUTONOMA(UAM).
- [9] MINSKY, M., (1968) *Semantic Information Processing*, MIT Press, Cambridge, Massachusetts.
- [10] POPOVICI, E.-C. (janvier 2008). *Information Retrieval of Text, Structure and Sequential Data in Heterogeneous XML Document Collections*. BRETAGNE.
- [11] Peng.W Dongqing Z ; Gang.Z ; Jingdong.W (2012) *Contextual Dominant Color Name Extraction for Web Image Search* . Peking Univ., Beijing, China
- [12] RIJSBERGEN, M. D. (May 1993). *HYPERMEDIA AND FREE TEXT RETRIEVAL*. University of Glasgow.
- [13] Schäuble,P. (1997). *Multimedia Information Retrieval: Content-Based Information Retrieval from Large Text and Audio Databases*. Springer

- [14] Sundgren, B. (1973). *An Infological approach to Data Bases*. Doctoral thesis, University of Stockholm, Department of Administrative Information Processing. Central Bureau of Statistics, Sweden.
- [15] Tian, D. Q. (October 2007). *Multimedia Information Retrieval*. San Antonio.
- [16] Udell, J. (2005, last viewed Aug. 2006). *Managing Metadata..* InfoWorld Oct. 20, 2005.
- [17] WINOGRAD, T., (1972) *Understanding Natural Language*, Edinburgh University Press, Edinburgh.
- [18] <http://www.WorkWithcolor.com> consulted in 02/04/2013

## ABSTRACT:

The aim of this dissertation is to provide a Multimedia Document Search Engine Software that retrieve information from digital libraries which are expanding every day in an increasing rate. We explore different methods of retrieving multimedia objects and explaining the integration between them, we also bring up proposed dominant color algorithm trying to guarantee color feature relying on metadata for selecting the appropriate results to fit user need.

## RÉSUMÉ

Le but de ce mémoire est de fournir un moteur de recherche dans les documents multimédias qui recherche de l'information dans les bibliothèques numériques devenant aujourd'hui très populaire. On explore les différentes méthodes de recherche des objets multimédias et on explique comment ces méthodes interagissent entre elles. Nous proposons également un algorithme pour l'extraction de couleur dominante en se basant sur les métadonnées des objets pour déterminer le bon résultat à l'utilisateur.

خلاصة:

إن الهدف الرئيسي لهذه المذكرة هو إنشاء تطبيق محرك بحث من أجل الملفات متعددة الوسائط الذي جعل خصيصا للبحث في المكتبات الرقمية التي أصبحت اليوم جد منتشرة. سوف نستكشف الطرق المختلفة للبحث في كائنات متعددة الوسائط كما سوف نشرح كيفية العمل مع بعضهم البعض. سنقترح أيضا خوارزم من أجل استخراج اللون الغالب بالاعتماد على البيانات التعريفية للكائنات لأجل تحديد نتائج مرضية للمستخدم.