University of Fort Hare
Together in Excellence

Development of a MPEG-7 based Multimedia Content Description and Retrieval Tool for Internet Protocol Television (IPTV)

By

Prince Daughin Ngqabutho Ncube

Supervisor

Dr. Zelalem Sintayehu Shibeshi

A thesis submitted in fulfillment of the requirements of the Degree of Master of Science in Computer Science at the University of Fort Hare

February 2017

Declaration

I hereby declare that “Development of a MPEG-7 based Multimedia Content Description and Retrieval Tool for Internet Protocol Television (IPTV)” is my original work and it has not been submitted before for any degree or examination at any other University. All sources I have used, consulted or quoted are duly indicated and acknowledged herein.

……………………………..

February 2017

Publications

Part of the research work presented in this thesis has been published in the following paper:

P. D. N. Ncube and Z. S. Shibeshi, Design of an MPEG-7 based Multimedia Content Description and Management System (MCDMS) for Internet Protocol Television (IPTV). In Proceedings of the 2015 Southern African Telecommunications Networks and Applications Conference, Arabella Hotel and Spa, Kogelberg Biosphere Reserve near Hermanus, Western Cape, South Africa.

Dedication

In loving memory of my mother

Acknowledgements

First and foremost I would like to thank the Lord God Almighty, whose nurturing hand continues to guide me through the capricious avenues of life. I am deeply indebted to my research supervisor Dr. Z. S. Shibeshi for his incalculable insight, inspiration, guidance, support, and the constructive criticism he gave me during the course of this research. Special thanks to the staff of the Department of Computer Science at UFH, to the H.O.D Mr. M. S. Scott, thank you for accepting me for the masters program and granting me the required equipment. I would like to acknowledge the support granted to me by the entire masters class during my study. To my family, I express my sincere gratitude for the unwavering love and support you have given me, needless to say, I love you guys. To the love of my life, Trezah, thank you for sticking with me through all my endeavors, I would not have succeeded without your support.

This work is based on research undertaken within the Telkom CoE in ICTD supported in part by Telkom SA, Tellabs, Saab Grintek Technologies, Easttel, Khula Holdings, THRIP and National Research Foundation of South Africa (UID: 90434). The opinions, findings and conclusions or recommendations expressed here are those of the authors and none of the above-mentioned sponsors accepts any liability whatsoever in this regard.

Abstract

Search and retrieval of multimedia content from open platforms such as the Internet and IPTV platforms has long been found to be hugely inefficient. It has been noted that a major cause of such inefficient results is the improper labeling or incomplete description of multimedia content by its creators. The lack of adequate description of video content by the proper annotation of video content with the relevant metadata leads to poor search and retrieval yields. The creation of such metadata itself is a major problem as there are various metadata description standards which users could employ. On the other hand there are tools such as FFprobe that can retrieve important features of video that can be used in searching and retrieval. The combination of such tools and metadata description standards could be the solution to the metadata problem. The Multimedia Content Description Interface (MPEG-7) is an example of a metadata description standard. It has been adopted by TISPAN for the description of IPTV multimedia content. The MPEG-7 standard is rather complex, seeing as it has over 1200 global Descriptors and Description Schemes which a user would have to know in order to implement such technology. This complexity is a nuisance when we consider the existence of multitudes of amateur video producers. These multimedia content creators have no idea how to use the MPEG-7 standard to annotate their creations with metadata. Consequently, overloading of the IPTV platform with content that has not been annotated in a standardized manner occurs, making search and retrieval of the multimedia content (videos, in this instance) inefficient. Therefore, it was imperative to try and determine whether the use of the MPEG-7 standard could be made much easier by creating a tool that is MPEG-7 enabled which will allow for the annotation of video content by any user without concerning themselves about how to use the MPEG-7 standard. In attempting to develop a tool for metadata generation, it was incumbent for us to understand the

issues associated with metadata generation for users wishing to create IPTV services. An extensive literature review on IPTV standardization was carried out to determine the issues associated with metadata generation for IPTV and their proposed solutions. An experimental research approach was taken in an attempt to figure out if our proposed solution to the lack of technical expertise by users about the MPEG-7 standard could be the final solution to the metadata generation problem. We developed a Multimedia Content Description and Management System (MCDMS) prototype which enabled us to describe video content by annotating it with 16 different metadata elements and storing the descriptions in XML MPEG-7 format. Incremental development and re-use oriented development were used during the development phase of this research. The MCDMS underwent functional testing; smoke testing of the individual system components and Big Bang integration testing for the combined components. Our results indicate that the more metadata is appended to a video as description the better it is to search for and retrieve. The MCDMS hides the complexity of MPEG-7 metadata creation from the users. With the effortless creation of MPEG-7 based metadata, it becomes easier to annotate videos. Consequently, search and retrieval of video content becomes more efficient. It is important to note that the description of multimedia content remains a complex feat. Even with the metadata elements laid out for users, there still exist other issues that affect metadata creation such as polysemy and the semantic gap. However, the provision of a tool that does the MPEG-7 standardizing behind the scenes for users when they upload a video makes the description of multimedia content in a standardized manner a much easier feat to achieve.

Table of Contents

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Project
3GPP2	3rd Generation Partnership Project 2
A/V	Audiovisual
AS	Application Server
ATIS	Alliance for Telecom Industry
BC	Broadcasting
CATV	Cable Television
CDMA	Code Division Multiple Access
CP	Content Provider
DB	Database
DOS	Denial of Service
DS	Description Schemes
DTDs	Document Type Definitions
ECHO	European Cultural Heritage Online
EDGE	Enhanced Data for Global Evolution
EPG	Electronic Program Guide
ETSI	European Telecom Standards Institute
GSM	Global System for Mobile communications
GUI	Graphical User Interface
IBM	International Business Machines Corporation
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
IMS	Internet Protocol Multimedia Subsystem
IP	Internet Protocol
IPTV	Internet Protocol Television

ISO	International Organization for Standardization
ITU	International Telecommunication Union
ITU-T	International Telecommunication Union Telecommunication Standardization Sector
LAN	Local Area Network
MAN	Metropolitan Area Network
MCDMS	Multimedia Content Description and Management System
MCF	Media Control Functions
MDF	Media Delivery Functions
MMS	Multimedia Messaging Service
MPEG	Moving Pictures Experts Group
MPEG-7	Multimedia Content Description Interface
MRAS	Microsoft Research Annotation System
MRF	Media Resource Function
NASS	Network Attachment Subsystem
NGN	Next Generation Network
NISO	National Information Standards Organization
PSTN	Public Switched Telephone Network
QFME	Query and Feedback Management Engine
QoE	Quality of Experience
QoS	Quality of Service
RACS	Resource Admission Control Subsystem
RDF	Resource Description Framework
SA	South Africa
SCF	Session Control Functions
SDF	Service Discovery Functions

SDO	Standard Development Organization
SDP	Session Description Protocol
SGML	Standard Generalized Markup Language
SMS	Short Message Service
SIP	Session Initialization Protocol
SMAT	Synchronous Multimedia and Annotation Tool
SSF	Service Selection Function
TISPAN	Telecoms and Internet Converged Services and Protocols for Advanced Networks
TV	Television
UCT	University of Cape Town
UE	User Equipment
UGC	User Generated Content
UMTS	Universal Mobile Telecommunications System
UPSF	User Profile Server Function
VD	Video Database
VIZARD	Video Wizard
VoD	Video on Demand
WAN	Wide Area Network
Wi-Fi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WTSA	World Telecommunications Standards Assembly
WWW	World-Wide-Web
XML	Extensible Markup Language

1. Chapter 1: Introduction

This chapter serves as an introduction to the research and to the chapters that follow. A comprehensive background into the research is given first. Thereafter, the problem statement and the research goals that guide the investigation of the research are provided. This is followed by a brief section that presents and clarifies the scope of research detailing the methodology used for investigating the research problem highlighted in the thesis. The final section is devoted to the description of the thesis structure and conclusion of the chapter.

1.1 Research Background

A lot of technicalities arise when trying to define multimedia. Computer scientists use the term "multimedia" to refer to anything that is not conventional alphanumeric data (Vries, 1999). Oftentimes, the term is made more explicit by an enumeration of data types. This enumeration of data types has led to the formulation of numerous definitions with regards to multimedia. An example would be, "Multimedia is the integration of multiple forms of media; this includes text, graphics, audio, video, etc." (Vries, 1999). The definition of multimedia more precisely than by a list of media types has always been a difficult task (Marcos, Illarramendi, Olaizola and Flórez, 2011). Therefore, a typical example of multimedia would be a presentation involving a combination of any of the above-mentioned data types, say, for example, a compilation involving audio and video clips. It comes as no surprise that such multimedia (video, audio, etc.) is used globally.

People interact with multimedia data daily, reading books, watching television, listening to music, etc. seeing as human beings make use of their five senses each and every day. Multimedia is closely related to the way humans experience the world daily. The very fabric of human communication is embodied in multimedia. Communication enables us to share experiences and create, maintain, sustain, and propagate knowledge. Mankind's innate desire for communication of experiences across time and space repeatedly, without having to be at the exact location of the person they are imparting the experience to, and the desire to impart this knowledge to an abundant number of individuals have been the foundations of multimedia (Friedland and Jain, 2014). This intrinsic desire resulted in several influential inventions that determined the progress of human civilization.

Human civilization has progressed since the invention of spoken languages, written languages, paper and the subsequent invention of Gutenberg's movable printing press to the invention of the telegraph, the telephone, the radio, movies, and television. Moreover, human progress did not stop there, the invention of recording media ensued, digital media became a reality, and the Internet surfaced. These technologies all brought with them room for the emergence of multimedia. This progress aimed to fulfill the need to cut across time and space and eventually led to the evolution of computing and communication technology. However, this progress came with by-products e.g. an exponential growth in multimedia content that needs to be organized. There is a dire need to organize and structure this multimedia in such a way that it can be easily accessed again when required. It is human nature to create photo albums, store compact discs with our music or, even have heaped boxes of our old newspapers, magazines, and files. This multimedia content becomes nugatory if it cannot be retrieved when searched for. Therefore, digitization emerged. Digitization brought with it better ways of sharing and storing multimedia. Multimedia digital libraries emerged to be the dynamo of multimedia data distribution and storage as they offered effective interaction among knowledge producers, librarians, and information and knowledge seekers (Adam et al., 1996).

The development of Internet technologies and digitization has given birth to the exponential growth of multimedia content on the World-Wide-Web (Perperis, Tsekeridou, and Theodoridis, 2007). With the recent trends in technology, it is easier to create, share and retrieve multimedia content from the Internet (Spaniol, Klamma, and Lux, 2008). YouTube, which is regarded as the world's largest online video sharing site, has been recorded as receiving a staggering 300 hours of video uploaded per minute (YouTube, 2015). However, retrieval of the multimedia content has become rather cumbersome with the increasing amount of content uploaded on the Internet and inadequate search results are often yielded. Information overload and the improper labeling/tagging of the multimedia content have been noted as some of the key causes of such inaccurate search results. The massive explosion in the variety of multimedia content on offer to date has become a reality since almost everyone nowadays has a camera, a digital printer or another device that almost instantly generates multimedia content. In most cases, the content creator wants to share their masterpiece with the world of multimedia consumers. This sort of behavior is sure to lead to an infinite amount of multimedia uploads on the digital repositories which are often linked to the world-wide-web.

Technology advancement surely comes with its setbacks. Access to the Internet is open to multitudes of people who upload as much multimedia content as they can afford. This increases the number of incompletely described multimedia content uploaded on the web which in turn puts pressure on consumers of multimedia content when they search for a particular content. The existence of projects like Project Isizwe: Tshwane Free WiFi puts more pressure on multimedia content description and retrieval. Project Isizwe is a non-profit organization which aims to bring the Internet to people across South Africa, by facilitating the roll-out of Free WiFi for public spaces in low-income communities, with a core focus on connectedness for the purpose of education. Project Isizwe believes that in today's connected world, access to the Internet should be considered an essential service, like water and electricity. Like water and electricity, it should be available to everyone, regardless of circumstance. They guarantee free WiFi access with no passwords or logins in public spaces. The rapid increase in WiFi-compatible devices throughout the country has allowed WiFi to become the most commonplace form of connectivity after 3G. WiFi thereafter snubs 3G when it comes to cost effectiveness in rural areas when we consider the prohibitive costs of 3G in rural areas. Therefore WiFi is the most appropriate access medium for rolling out networks offering the best-effort connectivity. With these Free Internet Zones (FIZs) the use of mobile devices is bound to escalate exponentially and with this escalation, an increase in creation and consumption of multimedia content is sure to ensue. Figure 1-1 and Figure 1-2 highlight the Internet traffic encountered in Tshwane alone, as a result of the free WiFi campaign.

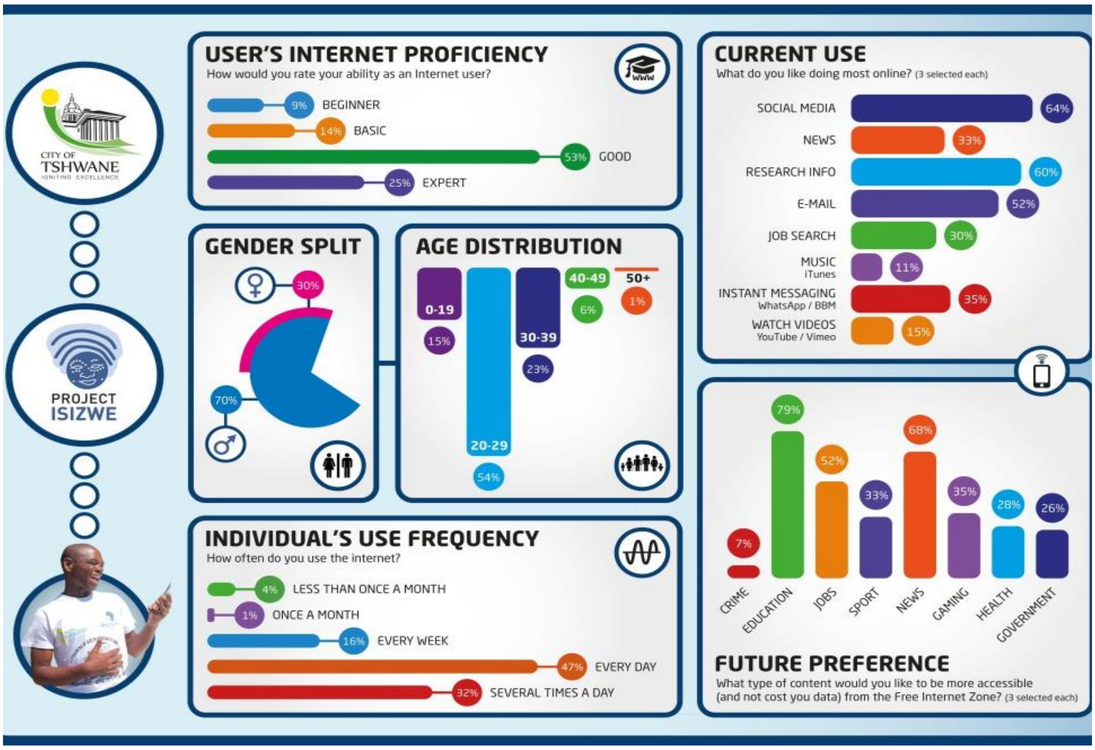


Figure 1-1: Project Isizwe Statistics (Source: (Project Isizwe, 2015))

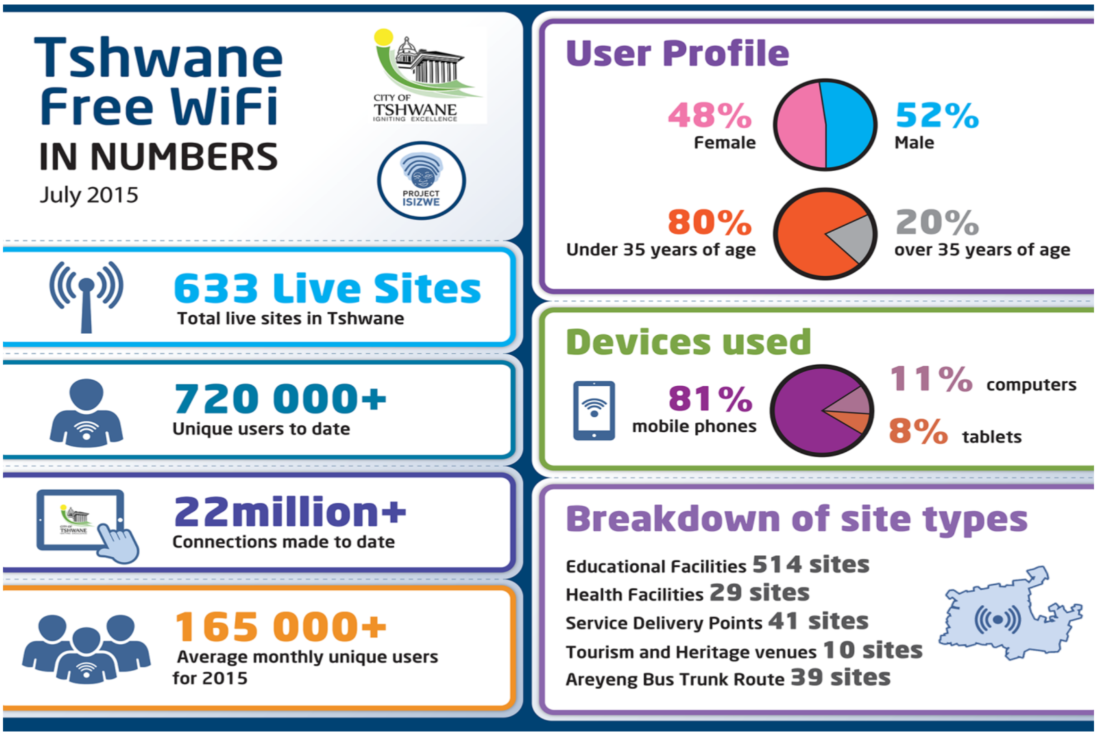


Figure 1-2: Project Isizwe Infographic (Source: (Project Isizwe, 2015))

1.2 Problem Statement

One of the major problems that people face on the Internet is information overload. There is too much information for any query we make and we get overwhelmed by the response we receive from the retrieval system, like Google. Accessing audio and video used to be a simple matter, simply because of the simplicity of the access mechanisms and because of the scarcity of the sources. Contrary to that, an abundant amount of audiovisual information has become available in digital form, in digital archives, on the World-Wide-Web, in broadcast data streams and in personal and professional databases, and this amount continues to grow, thereby worsening the search and retrieval process (José M. Martínez (UAM-EPS-GTI, 2004). The problem encountered from the information overload is magnified when it comes to multimedia data where direct searching of multimedia is in its infancy. We still use text to search for multimedia items that we are interested in and unless the content is properly described using appropriate metadata information, we cannot find it.

Different solutions are proposed to curb this problem and one solution is the proper tagging or labeling of digital media items using metadata. Metadata is created for multimedia items and the search is made based on the metadata. There are different standards developed to create and document metadata for multimedia items. One popular standard is MPEG-7, which is developed by the International Organization for Standardization (ISO). Telecoms and Internet Converged Services and Protocols for Advanced Networks (TISPAN), the major Internet Protocol Television (IPTV) standard body, also adopted MPEG-7 as a metadata description language. Considering the fact that, the value of information often depends on how easy it can be found, retrieved, accessed, filtered and managed; this project aims at developing a tool that can be used to easily describe and retrieve multimedia content that will be delivered in the IPTV system using the MPEG-7 standard.

1.3 Research Goals

Multimedia content description and retrieval is largely dependent on metadata. It is the metadata associated with the multimedia content that we term "description" and it is the same metadata we use when we search for and retrieve the content. For this research, it is essential that we undertake an investigation into the various complexities of the so-called "metadata".

Therefore, the first objective of the research is:

- To investigate issues related to metadata generation and retrieval.

IPTV is a platform to deliver innovative services. It is open and as such, it is safe to assume that whatever service we get from the Internet we should be able to get from IPTV. However, that assumption is not truly founded.

Therefore, the second objective of the research is:

- To investigate the important services of IPTV and the innovative services that could be created for IPTV consumption.

The third objective of the research is:

- To develop a MPEG-7 based multimedia content description and retrieval tool.

The fourth objective of the research is:

- To conduct functional testing of the tool developed in this research.

1.4 Contribution of study

It is no secret that the world has embraced the IPTV technology as one of the major video services of the telecommunications industry. The world all over is moving towards digitization. The United Kingdom (UK) completed its digital switchover in the year 2012 (Jambor, 2012). South Africa (SA) itself specified a roll out stage towards digitization during the World Telecommunications Standards Assembly (WTSA) conference in 2008. With the global switch to digitization, IPTV becomes a perfect platform for providing converged services. IPTV brings with it the User Generated Content (UGC) service. This means any user (who may not have the expertise of metadata creation) can generate content and upload it as UGC; as a result, it is very useful to have a tool that will assist users in describing their content for the purpose of multimedia content search and retrieval. Therefore, the major contributions of this research are:

- The open source tool, that can be used to describe multimedia content in a standard way that enhances multimedia search and acquisition, is the major contribution of this research. This tool will help both the novice users and big broadcasting companies alike. It may also be used in the research community to develop other services that require metadata generation and retrieval. Therefore, this tool will help aid the rollout of IPTV in SA.
- The theoretical contribution related to metadata creation and retrieval, in particular,

that of IPTV.

1.5 Scope of the research

To achieve the goals mentioned in 1.3, the following methodology is employed:

- A technical study on multimedia content descriptors, particularly in the domain of metadata creation and retrieval. The extensive literature review on multimedia content descriptors, metadata generation, acquisition, and representation is undertaken to provide background knowledge necessary for understanding and developing a multimedia content management system, particularly for IPTV. We hypothesize that a MPEG-7 based Multimedia Content Description and Retrieval System/Tool could be the solution to curbing the cumbersome search and retrieval of multimedia content that has not been described in a standard manner (such content lacks metadata that can be easily retrieved when searching for the content).
- Incremental prototyping is used during the developmental stage of the research. A required functionality is implemented then tested for any bugs. If and only if the functionality proves bug free is another functional unit added on top of the current prototype.
- The coding and implementation phase of the research are done using the C++ programming language. The object oriented C++ is chosen because it is a common programming language that can be adopted by future enthusiasts of this research space.
- The developed prototype is evaluated based on the hypothesis given. We test to see if the tool can describe the multimedia content and encode it in a MPEG-7 format so that it can be retrieved later by another MPEG-7 enabled application.

However, it is vital to note that in this particular research no focus was given on how IPTV handles access, connectivity, subscription accounts, etc. The entire research was focused on Multimedia Content Description and Retrieval.

1.6 Overview of the Dissertation

To achieve the aims and objectives defined above, the thesis consists of six chapters, including this introductory chapter.

Chapter 1: introduces the research.

Chapter 2: gives a detailed literature review.

Chapter 3: presents on the methodology.

Chapter 4: dwells on the design and implementation.

Chapter 5: describes the testing, results, and presents the discussions of the research.

Chapter 6: presents the conclusions drawn from the research. It goes further to highlight areas that can be explored as future work in this research space.

1.7 Summary

This chapter dwelt on introducing the research study by presenting a background to the study, stating the problem and goals of the research. The contribution of the research was highlighted before addressing the methodology employed to achieve the aforementioned research goals. The chapter concluded with the structure of the thesis. The ensuing chapter gives a detailed literature review on multimedia content description and retrieval in IPTV.

2. Chapter 2: Background and Literature Review

This chapter provides an in-depth review of multimedia content description. It describes the current state-of-the-art networks and tools used for multimedia content streaming. The standards associated with multimedia content description are described with an emphasis given to MPEG-7. This chapter concludes by giving a review of MPEG-7 based systems as related work to this research.

2.1 State-of-the-art-networks and tools

2.1.1 Next Generation Network (NGN)

A number of Standard Development Organizations (SDOs) have contributed to the description of NGN. The key players in NGN standardization are the International Telecommunication Union Telecommunication (ITU-T) Standardization Sector, the Alliance for Telecom Industry (ATIS), the European Telecom Standards Institute (ETSI) and the 3rd Generation Partnership Project (3GPP). The ITU-T defines NGN as a "*packet based network able to provide services including telecommunication services and able to make use of multiple broadbands, QoS-enabled transport technologies and in which service related functions are independent of underlying transport-related technologies*" (ITU-T, 2015). Regardless of which SDO defines NGN, their definitions share common characteristics, namely:

- Unified Global Networking Platform
- Packet-Based Network
- Provides Telecommunication Services to Users
- QoS-enabled Transport Technologies
- Services are Independent of Transport Technologies
- Unfettered Access for Users to Networks and Services

- Generalized Mobility Which Allows Consistent and Ubiquitous Provision of Services to Users

  To date, common networks are composed of the Public Switched Telephone Network (PSTN), the Packet Switched Networks (LAN, WAN, MAN, etc.), and Cellular Mobile Networks. The goal of the NGN is to bring together different and provide convergence. Convergence refers to the shift from the traditional "vertical silos" architecture, *i.e.* a situation in which different services were provided through separate networks (mobile, fixed, cable television, IP), to a situation in which communication services will be accessed and used seamlessly across different networks and provided over multiple platforms, in an interactive way (Sarrocco & Ypsilanti 2011). It is the process of interconnecting the traditional switched circuit networks (PSTN and Mobile Networks) and packet-based networks that use the IP. The whole philosophical basis of NGN is to unify data, telephone, and mobile networks together with their services into one network concept. An illustration of the shift from the traditional vertical silo architecture to an NGN setup is shown in Figure 2-1. The figure shows the differences between telecommunication services that were developed in a stovepipe fashion ("vertical silo"), specifically attached to a special access network and those developed in an NGN multi-service architecture. Before the introduction of NGN, a user who had a voice mail from his mobile network provider could only access his voicemail using his cell phone and no other device could be used. But, as time went on, consumers of telecommunication services wanted to be able to access their varied services via different devices without having to worry about the device they were using. This then led to the need to converge the various networks, and thus emerged the NGN.

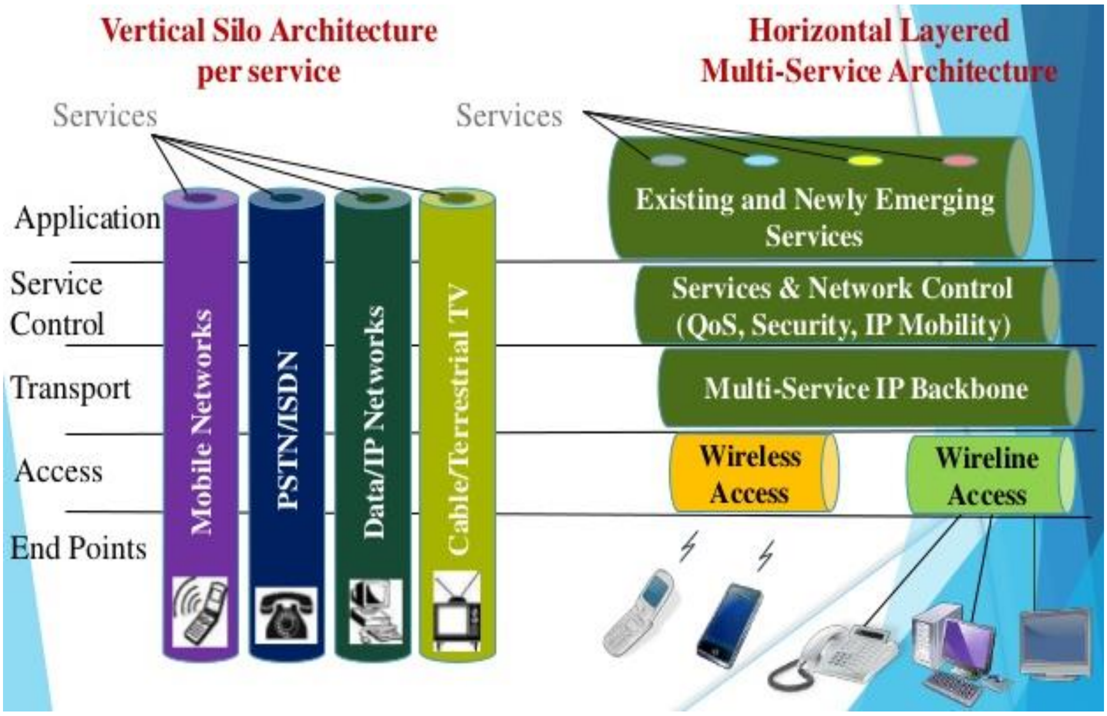


Figure 2-1: Shift towards Convergence (Source: (Bathaee, 2014))

NGN offered a simplified service model to service developers where they can concentrate on developing their services regardless of the access media that will be used to access that particular service. We, therefore, observe the ninety-degree shift of the vertical silos with each independent service to give a horizontal layered multi-service architecture as depicted in Figure 2-1. Figure 2-2 highlights how the redundant dedicated technologies with duplicated functions in the stove pipe setup are then transformed into a converged setup when the NGN is employed offering a much simpler and compact service architecture.

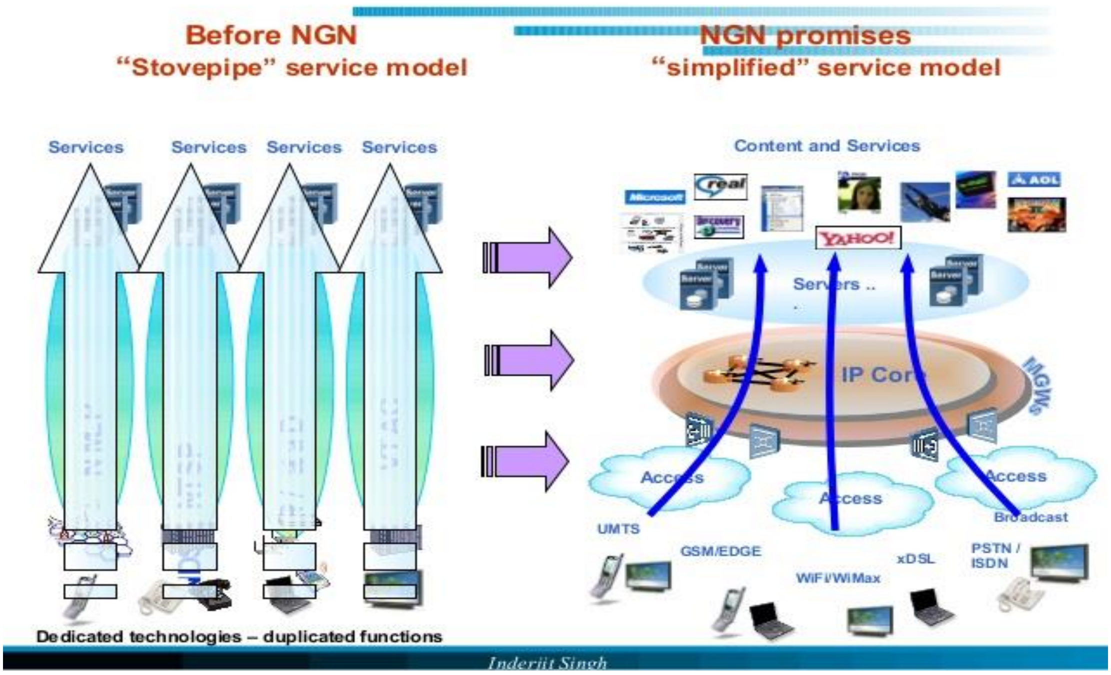


Figure 2-2: Convergence in NGNs (Source: (Singh, 2015))

2.1.2 IP Multimedia Subsystem (IMS)

The authors (Tsai et al., 2010) describe IMS as a service framework and a core network which provides integrated IP multimedia services that are based on Session Initialization Protocol (SIP) to support service session control functions. According to (Lindgren, 2010), with the vision of creating a new generation of cellular networks using IP-based communication, service providers, operators and developers have created the IMS. The author goes on to state that *the Third Generation Partnership Project (3GPP) introduced the IP Multimedia Subsystem (IMS) to combine the cellular communication networks with Internet technology in order to facilitate the ability of developers to introduce new services*. The 3GPP is a consortium of global standards organizations that developed the IMS standardization document. Owing to the credibility and popularity of the 3GPP, the IMS has been adopted by several international standards bodies, including TISPAN.

The IMS is an IP multimedia and telephony core network service subsystem within the NGN architecture that has been defined by the 3rd Generation Partnership Project (3GPP) and the 3rd Generation Partnership Project 2 (3GPP2) in accordance with the Internet Engineering Task Force (IETF) Internet protocols. It is access independent as it supports IP to IP session over wireline IP, Wireless Fidelity (Wi-Fi), Worldwide Interoperability for Microwave Access (WiMAX), Code Division Multiple Access (CDMA), packet data along with Global System for Mobile communications (GSM)/Enhanced Data for Global Evolution (EDGE)/Universal Mobile Telecommunications System (UMTS) and other packet data applications (Beal, 2015). The IMS was introduced by the 3GPP initiative as the architectural subsystem dedicated to controlling and providing multimedia services over packet-based core network within third generation mobile networks (Mikoczy, Sivchenko, Xu and Rakocevic, 2007).

The IMS supports conversational services such as voice and video telephony, push-to-talk/show/share, short message service (SMS) and multimedia messaging service (MMS) services. It now becomes the core of NGN to provide multimedia services (ETSI Technical Committee TISPAN 2009). With such a framework as the IMS, it then opens up an avenue for interactive TV and as such, we have IPTV now coming into the picture.

2.1.3 IP Television (IPTV)

IPTV is defined by the International Telecommunication Union (ITU) focus group on IPTV as multimedia services such as television/video/audio/text/graphics/data delivered over IP based networks that are managed to provide the required level of service quality and experience, security, interactivity, and reliability (ITU-T Newslog, 2006). The Alliance for Telecommunications Industry Solutions (ATIS)

IPTV Exploratory group, on the other hand, defines IPTV as the secure and reliable delivery to subscribers of entertainment video and related services (e.g. Live TV, Video on Demand, Personal Video Recorder, etc.) (ATIS IPTV Research Group, 2005).

There has always been a lot of confusion between Internet TV and IPTV. It should be noted that while Internet TV travels over the open, public, global Internet, IPTV uses a private, managed network. IPTV technology allows television services to be delivered over a proprietary broadband packet data network using the internet protocol suite. Internet TV is a television broadcast service distributed over the Internet. Oftentimes, IPTV has been dubbed as a platform to deliver innovative services. It is open and allows for the creation and deployment of innovative services by subscribed consumers. IPTV has the ability to deliver a limitless number of services; the major services on offer are as follows (ETSI Technical Committee TISPAN, 2009):

- Broadcast TV (with or without trick modes).
- Content on Demand (CoD).
- Personal Video Recording (cPVR, nPVR).
- Pay Per View (PPV).
- Interactive TV (iTV).
- near CoD (nCoD).
- push CoD (pCoD).
- User Generated Content (UGC).
- Profiling and personalization.
- Content Recommendations (CR).
- Advertising (Ad) and Targeted Advertising (TAI).

- Messaging services.
- Notification services.
- Personalized channel.
- Bookmarks or Content Marking (CM).

User Generated Content (UGC) as the name implies, is content generated by a user for the purpose of making it available to other users. The content may be for a specific audience or public for everyone to consume. This type of content is of much interest to us as it entails the creation of multimedia content by users, who will then upload such content onto the IPTV platform for consumption by their peers. UGC as enticing as it sounds could be a nightmare when we think about content retrieval. The users who create such content are amateurs in multimedia content description, which means they lack the required expertise to label the content with its relevant metadata. Without its relevant metadata associated with it, multimedia content is next to impossible to search for and acquire. This then worsens the information overload that we already suffer from.

IPTV brings with it an open platform for the creation and deployment of innovative services, and we should be able to benefit from the services on offer. However, to consume these services we should first be able to retrieve the content they deliver. Retrieval of this multimedia content requires proper description and annotation of the created multimedia content. To date, we do not have an open source tool that can help users describe UGC in a standardized manner to ensure the multimedia content can be retrieved as effortlessly as possible when searched for. The fact that users do not know how to annotate the multimedia content they create is just but the beginning of a problem. Even when experts are used when describing the multimedia content created for IPTV, the workload would be absurd as we account for the vast amounts of services created and deployed. As stated above, IPTV has the ability to deliver an abundant number of value-added services. Try to imagine a situation whereby a certain number of experts try to annotate multimedia content for the different innovative services listed below:

- Information browsing
- Interactive advertising
- Interactive games
- Interactive guessing games
- Online transactions
- TV Mail
- TV Banking
- Electronic Program Guide
- TV Billing
- Reminders
- Map Mash-ups
- TV Chat
- TV Album
- Caller ID
- T-Commerce
- Content Search
- Health and Fitness
- Society Information Service
- VoIP
- Music Downloads
- Security
- Teleworking
- File Transfer/Cloud
- Home Networks

All these value-added services can be created rapidly and uploaded onto the IPTV platform by various users at any given time. The metadata associated with such multimedia content requires a long list of descriptors for its annotation. This annotation is a cumbersome process when done manually, therefore an automatic or semi-automatic way of annotating such created content would go a long way in reducing the workload, therefore, ensuring efficient and adequate description of multimedia content for the various services created. This description will, therefore, quicken the acquisition of the multimedia content when searched for by the millions of IPTV subscribers globally.

2.1.4 Evolution of IPTV architecture

During the evolution of IPTV service architecture; there have been three dominant types of IPTV architectures that have been implemented. These three emanate from two groupings i.e. NGN-based and non-NGN-based IPTV architecture. The IPTV functional architecture in Figure 2-3 below can be used to represent all three types of IPTV architectures. The difference between them is further elaborated below.

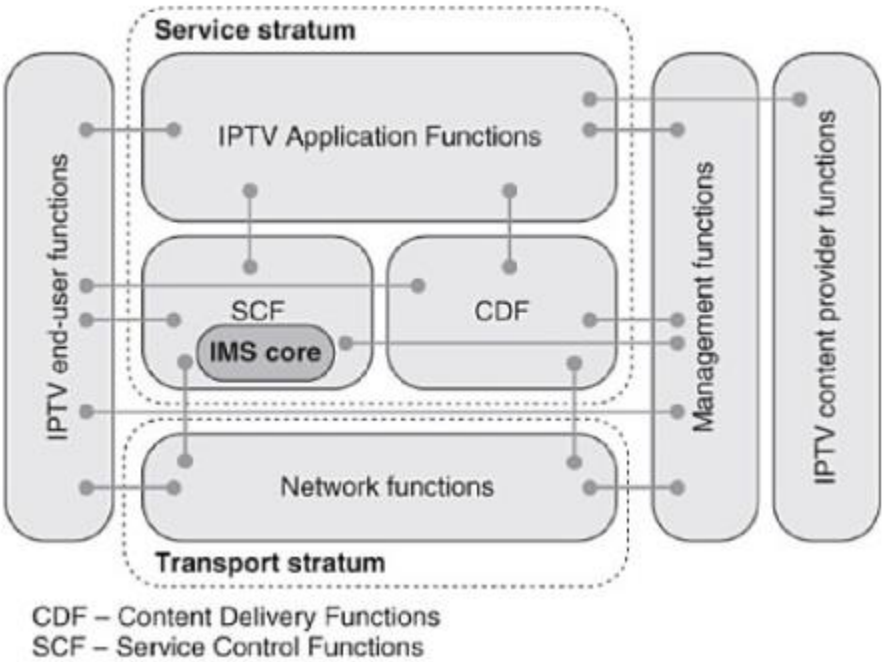


Figure 2-3: IPTV functional architecture based on IMS (Source: (Janevski, 2014))

2.1.4.1 Non-NGN based IPTV architecture – it was the earliest IPTV architecture and it was predominant for the period up till 2010.

2.1.4.2 NGN-based non-IMS IPTV architecture – it is also called NGN Integrated IPTV. It came during the transitional phase of IPTV, from non-NGN based IPTV and before IMS-based IPTV was implemented.

2.1.4.3 NGN-based IMS-based IPTV architecture – as the name implies, this architecture has both the NGN service components and the IMS with its functions located in the service stratum of the NGN. This type of architecture uses the IMS for controlling the IPTV service component. NGN-based IMS-based IPTV is more commonly referred to as IMS-based IPTV because there is no Non-NGN-based IMS-based IPTV architecture.

2.1.5 NGN-based IMS-based IPTV

From the beginning, IMS was introduced with the idea of being a multimedia delivery platform, however, the standard IMS architecture through its component called MRF (Media Resource Function) cannot provide streaming and broadcasting (BC) services; it just provides video-oriented services like video announcement and recording services. Therefore, this is the reason why TISPAN introduced the IMS-based IPTV architecture which would cater for services like BC and streaming services. Figure 2-4 depicts the IMS-based IPTV architecture.

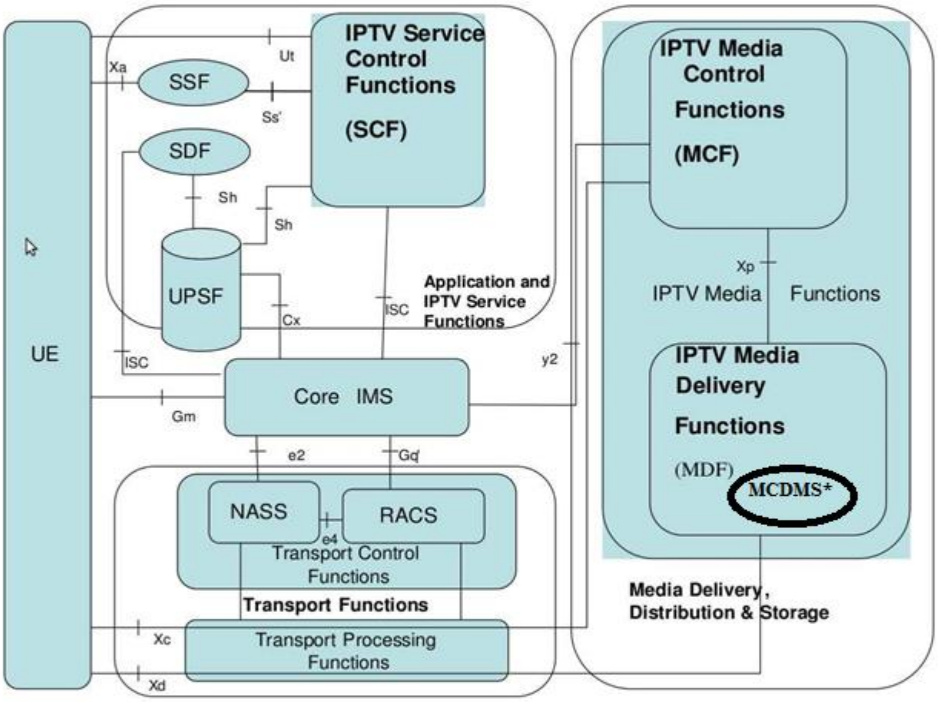


Figure 2-4: IMS-based IPTV architecture with our proposed MCDMS system embedded into it

User Equipment (UE) – this refers to the functional entity that provides the user with access to IPTV services.

Service Selection Function (SSF) - a functional entity that provides service selection information to the UE

Service Discovery Function (SDF) - a functional entity that provides service attachment information to the UE.

User Profiles Service Function (UPSF) – an entity that holds the IMS user profiles and possibly IPTV specific profile data.

Session Control Function (SCF) – the functional entity that provides IPTV service logic and the functions required to support execution of such logic

Media Control Function (MCF) - the functional entity that provides the UE with functions required to control media flows and manages the MDFs under its control.

Media Delivery Function (MDF) - the functional entity that delivers content data to the UE.

Multimedia Content Description and Management System (MCDMS) – this is the tool we developed in this work. It falls under the Media Delivery functions of the IMS-based IPTV architecture.

2.2 Multimedia Content Description

Describing content is a process in itself. It involves gathering essential characteristics about the content in question that may be used in a later process to more easily retrieve, filter or adapt that particular content. Therefore it is safe to conclude, that content description is intended to facilitate the access to the pretended content. The description of multimedia content can be referred to as metadata. This emanates from the fact that metadata is description. Description of what you might ask. The following section breaks it down:

2.2.1 What is metadata?

It stands to reason that the title and time stamp of a video are information about the video. They describe the video, and they are classified as metadata. Metadata is description of anything (physical/digital/natural/artificial). The ETSI TS 102 822-2 defines metadata as descriptive data about content, such as title and synopsis (ETSI Technical Committee TISPAN, 2007). The National Information Standards Organization (NISO) Press defines metadata as structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information source (National Information Standards Organization, 2004). This deduction of metadata then leads us to the various types of metadata i.e.

- Descriptive metadata e.g. title, author, publication year, etc. It describes a particular resource for purposes such as discovery and identification.

- Structural metadata e.g. compound objects – book, chapters, pages, sections, etc. It specifies how composite objects are put together.
- Administrative metadata e.g. do not bleach, access rights, copyrights, providence, etc. It provides information to aid in managing a particular resource.

In this research, we focused mostly on descriptive metadata, although we do include a bit of structural and administrative metadata elements as part of the 16 basic metadata elements required by the MCDMS for video description.

2.2.2 Metadata Creation

Who creates metadata is a question that still baffles a lot of individuals. An appropriate response to that question would be to state that, metadata creation varies by discipline, the resource being described, the tools available, and the expected outcome. However, metadata creation is oftentimes a joint effort (National Information Standards Organization, 2004). Most basic structural and administrative metadata is furnished by the technical personnel when the actual creation of the digital object is being done. It can also be imparted via an automated process following preset conditions by technicians. To achieve distinguished multimedia content descriptions, a variety of metadata creation tools can be used.

2.2.3 Metadata Creation Tools

There are four basic types of metadata creation tools. The tool chosen for a particular job depends on the requirements of the job in question. Various metadata creation tools can be combined at times to ensure optimum description of a resource is obtained. The different types of metadata creation tools are:

- Templates – these types of tools allows a user to enter the metadata values into preset fields that match the element set being used. The template then generates a formatted set of the element-value pairs.

- Markup tools – these tools structure the metadata attributes and values into the designated schema language. The majority of these markup tools generate Extensible Markup Language (XML) or Standard Generalized Markup Language (SGML) Document Type Definitions (DTDs). The templates defined above normally include such markup tools as part of their final translation of the metadata.
- Extraction tools – these tools automatically create metadata from an analysis of the digital resource in question. However, extraction tools are mostly limited to textual resources and the quality of the extracted metadata can vary significantly based on the tool's algorithms as well as the content and structure of the source text. It is for this reason that such tools are used as an aid in creating metadata. The metadata output of such tools normally requires manual revision and editing.
- Conversion tools – these types of tools translate metadata formats. The similarity of elements in the source and target formats affects the amount of additional editing and manual input of metadata that may be required.

It should be noted that metadata tools are generally developed to support specific metadata schemes or element sets. However, even with the abundance of metadata creation tools on offer, the creation of metadata is still a worrisome process. For starters, there is the need for metadata quality control. The automatic creation of metadata or metadata creation by information experts who are not familiar with cataloging, indexing, or vocabulary control can create quality problems. Mandatory elements may be missing or used incorrectly by amateur metadata creators. Metadata terminology may be inconsistent making it difficult to locate relevant information. It is for this reason that guidelines for good quality metadata creation are articulated (NISO Framework Working Group, 2007).

A lot of issues arise when referring to metadata. There is the metadata problem itself which includes:

- Vast amounts of audio, video, and other multimedia data
- The proliferation of TV and broadcasting stations
- Multimedia Content on the Internet (Real streams)
- Personal archives e.g. digital images, video, and audio

The goal is efficient archiving and managing of audiovisual (A/V) data, automatic methods to analyze multimedia content (recognition), standards to describe A/V content (metadata), and information about the content, speakers, transcripts, persons and time.

Unfortunately, the metadata problem is further enhanced when we consider the actual metadata creation or generation itself. Users are faced with complications with regards to complex metadata description standards, the semantic gap, the vocabulary problem, the encoding specificity problem, just to mention but a few. The complications are elaborated further as follows:

2.2.3.1 Metadata Description Standards

There exists an abundance of metadata description standards. Normally, each standard is developed for a specific purpose. Dublin Core (Dublin Core Initiative, 2013) was initially developed to describe Web resources, the Text Encoding Initiative (Text Encoding Initiative, 2016), for marking up electronic texts, Metadata Encoding and Transmission Standard (Library of Congress, 2016), for describing complex digital library objects and MPEG Multimedia Metadata Standard (The Moving Pictures Experts Group, 2016), for coded representations of digital audio and video. The numerous standards in existence all require some technical expertise. The standards vary in terms of relative advantage, compatibility, complexity/simplicity, trialability, and observability. These metrics affect the adoption of the various metadata description standards by content producers and prosumers alike. For example, Dublin Core is normally adopted for its simplicity. However, the complexity and compatibility of most of the existing metadata description standards have pressured service providers and users alike to evade the use of such standards. Consequently, this then leads to non-standardized metadata creation, thus complex multimedia content search.

2.2.3.2 The semantic gap

The semantic gap as defined by Snoek & Worring is "the lack of correspondence between the low-level features that machines extract from video and the high-level conceptual interpretations a human gives to the data in a given situation" (Snoek and Worring, 2007). The semantic gap is simply defined as the difference between the information that a machine can extract from the perceptual data and the interpretation that a user in a given situation has for the same data. It is this particular gap that has been a major scourge in information retrieval for decades. When retrieving multimedia content there are two aspects that can be considered, the low-level features like color, texture, shape, etc., and the high-level features like the text appended to the content. The high-level features contain the semantic value of the multimedia content which humans can handle quite well, but machines still struggle with semantics hence the difficulties when trying to fully automate metadata annotation. A typical example would be a scenario whereby we try to retrieve an image, the fundamental concern in the image retrieval is how to adopt/derive high-level concepts automatically from the image content and its surrounding text. The image content, unlike the text, has no clear semantic meaning. Analysis of multimedia content such as an image requires insightful thinking that computers are not capable of doing.

The text embedded in images itself involves some problems too. Primarily, if an image is embedded in a textual document or a web page, there is generally a lot of text, and the system has to decide which words are related to the image and which are not. Furthermore, even if words in a given textual document are related to the image content, subjectivity of using words in a given document for a given image can be a problem. Zhao and Grosky referred to this problem as synonymy and polysemy (Zhao and Grosky, 2002). Synonymy is used to describe the fact that there are many ways to refer to the same object, e.g., subject can also be issue, matter, case, business, course, etc. The prevalence of synonyms tends to decrease the fraction of relevant instances that are retrieved (recall performance). Polysemy refers to the fact that most words have more than one distinct meaning. For example, subject can mean theme and matter, but it can also be used in a court to force upon someone. Polysemy is a factor underlying poor precision performance. In an attempt to narrow the semantic gap many approaches have been developed. One approach is to combine both low-level features and textual features. A typical example of such an approach is Google's inside search (Google, 2017). Users can now begin their Google search with an image. To reduce the problems encountered due to subjectivity, automatic annotation of images could solve the problem. Latent Semantic Indexing (LSI) automatically indexes images with linguistic terms based on statistical model comparison. For example, images annotated by words "reservation", "book", "ticket" and "destination" are related to other images about transport companies e.g. buses, planes, etc. LSI approach, as a rule, is used together with the Content Based Information Retrieval (CBIR) approach to narrow the semantic gap.

2.2.3.3 The vocabulary problem

How often have you heard someone being called an individual? The very fact that people are different based on their DNA should be justification enough that their vocabulary could differ too. For example, one person may describe a picture of 'the beach during the day' as 'bright', whilst another person describes the same picture as 'vivid'. These different individuals try to describe the same picture, but if either of them tries to search for the picture collected by the other in the database, they both would fail to retrieve it. In the information retrieval domain, this is called the vocabulary problem. This is a common bane in information retrieval and it affects the production of metadata immensely. Attempts have been made to curb the vocabulary problem by making use of thesauri, semi-automatic query expansion, and developing standards.

2.2.3.4 The encoding specificity problem

It has been established above that people have differing vocabulary. This extends to people with different expertise. A picture portraying 'A Manchester Derby' by a football journalist may be associated with 'A match at the great Old Trafford' by an enthusiast architect. Most people may comprehend that as being highly likely, but they would not consider the fact that a single person will describe a single picture differently when asked about the picture at different times. The environment that a person is in influences the way they describe a particular art. For example, the Manchester derby could be described as 'mind-blowing' during the post-match interview, but be later described as 'a very technical match' after a week of analysis of the statistics of the game. This type of behavior is called the encoding specificity problem, and it is a nuisance in the metadata development industry (Vries, 1999).

2.3 MPEG-7

MPEG-7 standard, which is formally called "Multimedia Content Description Interface", is an ISO/IEC 15938 standard. It was developed by the Moving Pictures Experts Group (MPEG), the very same committee that developed MPEG-1, MPEG-2, and MPEG-4. However, unlike the other MPEG standards, MPEG-7 was designed to describe multimedia content. To achieve multimedia content description, MPEG-7 has specified tools attached to the standard. These tools are defined as follows:

- **Descriptors (D):** they represent a feature, and define both the syntax and semantics of a feature representation. Examples of descriptors are a color histogram, a motion field, the text of the title, etc.
- **Description Schemes (DS):** they specify both the structure and the semantics of the relationships between their components. These may be both Descriptors and Description Schemes. An example of DS could be a movie that has been temporally structured as scenes and shots, including some textual descriptors at the scene level, and color, motion, and audio descriptors at the shot level.
- **Description Definition Language (DDL):** allows the creation of new Description Schemes and possible Descriptors as well as the extension of existing Description Schemes.
- **System tools:** support the multiplexing of descriptions, synchronization of descriptions with the associated content, binary representation for efficient storage and transmission, management and protection of intellectual property, etc.

A particular multimedia content can be described using several Descriptors and description Schemes characterizing the content. MPEG-7 defines a minimum set of restrictions, which opens up new possibilities for applications to explore in terms of audiovisual description. It should be noted that new descriptors are already under study to be included in future versions of MPEG-7 in an attempt to provide additional description functionalities. Figure 2-5 below elaborates the organization of Descriptors and description Schemes when it comes to content description.

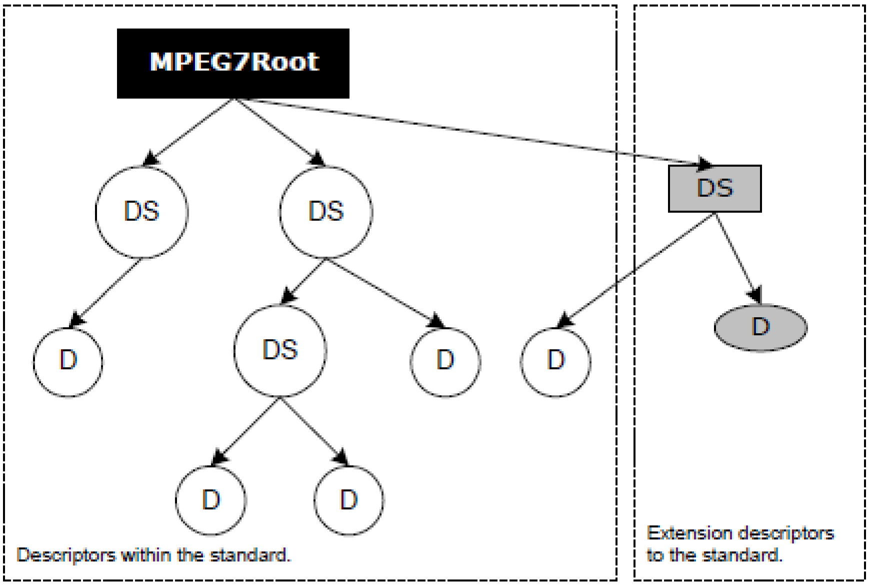


Figure 2-5: Relation between the different MPEG-7 description elements (Source: (Magalhaes, 2002))

As illustrated in Figure 2-5 above, a root element starts the tree describing a piece of multimedia content. Several Descriptors and Description Schemes can be used to describe the same piece of multimedia content. The fact that the MPEG-7 standard is extensible, allows for the creation of new Descriptors and DS as illustrated in Figure 2-5 above.

The next question that might arise is how are multimedia content descriptions generated and consumed? The block diagram in Figure 2-6 below gives a breakdown of how that is done. After extraction or human annotation of the features, the features are coded in the MPEG-7 format so that they may be consumed later by any MPEG-7 enabled application. The procedure goes as follows:

- Content is analyzed and features extracted (1) to generate a content description (2).

- The content description is then coded using the MPEG-7 standard on a MPEG-7 enabled framework.
- The description can be linked to the content it described (4) if desired.
- Using a framework that is MPEG-7 enabled, an application decodes the MPEG-7 standard content description (5).
- An application uses the MPEG-7 content description to easily retrieve, filter, manage, etc. the content (6).

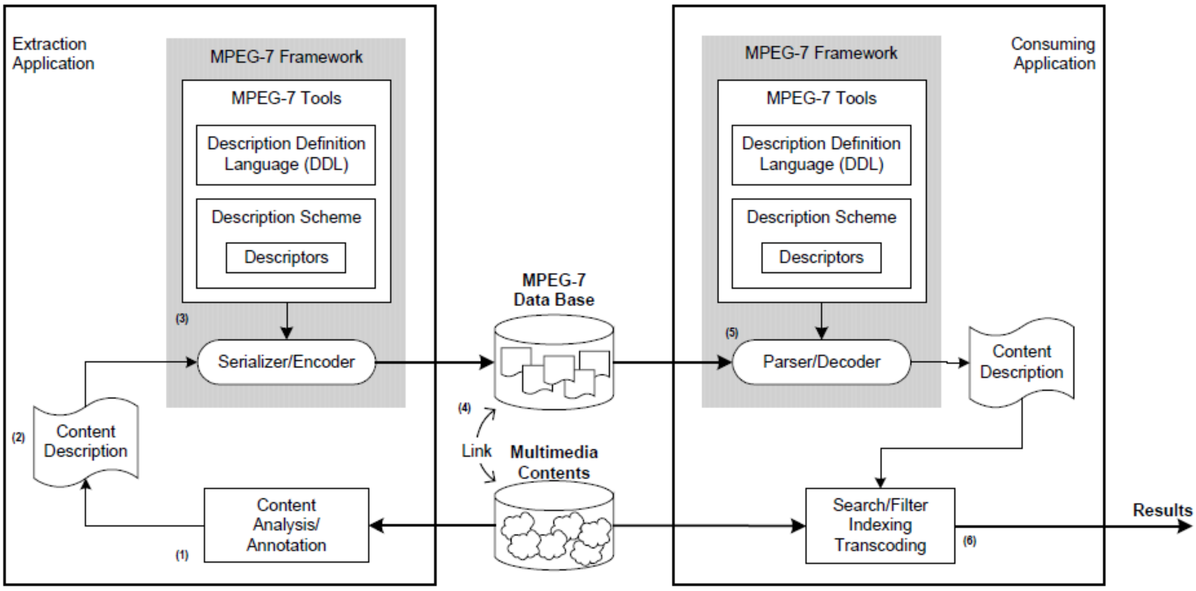


Figure 2-6: Abstract representation of applications using MPEG-7 (Source: (Magalhaes, 2002))

2.3.1 The MPEG-7 Standard Overview

The MPEG-7 standard is organized into eight parts, namely:

- **Part 1: MPEG-7 Systems:** includes tools that are required for the preparation of MPEG-7 descriptions for efficient transport and storage, and to allow synchronization between content and descriptions as well as the tools related to managing and protecting intellectual property.

The MPEG-7 systems architecture defines the basic structure through which MPEG-7 descriptions must be exchanged. Descriptions can be exchanged both in a textual format using the DDL language specified in **Part 2: MPEG-7 DDL** or in binary format using the compression tool specified in **Part 1: MPEG-7 Systems**. The main elements of MPEG-7 Systems are:

- **The terminal architecture**
- **The access units**
- **The normative interfaces**

- **Part 2: MPEG-7 Description Definition Language:** this component includes the languages for defining new Description Schemes and eventually also new Descriptors.

  The MPEG-7 Description Definition Language forms a core part of the MPEG-standard since it provides the solid descriptive foundation by which standard Descriptors and DS are defined. It allows users to create their own Descriptors. The DDL is able to express spatial, temporal, structural, and conceptual relationships between the elements of a DS, and between the DSs. The DDL provides a rich model for links and references between one or more descriptions and the data that they describe. It is platform and application independent and both human and machine readable. The DDL makes use of an extended version of the XML framework as agreed upon at the 51st MPEG Meeting held in Noordwijkerhout in March 2000 (Pereira, 2000).

- **Part 3: MPEG-7 Visual:** includes the Descriptors and Description Schemes dealing only with visual data.

  The visual description tools included in **Part 3: MPEG-7 Visual** consists of basic structures and descriptors that cover visual features in the terms of color, texture, shape, and motion. Basically, there are five basic structures used by MPEG-7 visual descriptors, namely:

- **Spatial 2D coordinates**
- **Temporal interpolation**
- **Grid layout**
- **Time series**
- **Multiple view**

The MPEG-7 standard has numerous types of descriptors; we will mention just five different types, namely (Magalhaes, 2002):

- **Color Descriptors -** there are seven color descriptors defined: color space, color quantization, dominant color, scalar color, color layout, color structure, and GoF/GoP color.
- **Texture Descriptors** – there are three MPEG-7 texture descriptors defined: homogenous texture, texture browsing, and edge histogram.
- **Shape Descriptors** – there are three shape descriptors defined: region-based, contour-based, and 3D shape.
- **Motion Descriptors** – there are four motion descriptors defined: camera motion, motion trajectory, parametric motion, and motion activity.
- **Localization** – there are two MPEG-7 localization descriptors defined: region locator, and spatial temporal locator.

- **Part 4: MPEG-7 Audio:** includes the Descriptors and Description Schemes dealing only with audio data.

Part 4 specifies description tools for audio data. The audio description tools are applicable to all forms of audio (music, speech, sound effects, soundtracks, and any mixtures of these). The MPEG-7 Audio part considers two major components, namely:

- Audio Framework: a collection of tools and low-level descriptors intended as a framework for the construction of many applications.
- High-level tools: a set of tools focusing on more specific applications, these tools may make use of the audio framework tools such as timbre description, sound recognition, spoken content, melody contour, melody, etc.

- **Part 5: MPEG-7 Multimedia Description Schemes:** includes the Descriptors and Description Schemes dealing with generic features and multimedia data.

  The MPEG-7 DSs expand on the MPEG-7 Descriptors by combining individual Descriptors as well as other DSs. Typically, the Multimedia Description Schemes (MDS) refer to all kinds of multimedia consisting of audio, visual and textual data, whereas the domain specific Ds or DSs refer specifically to the audio or visual domains. The MDS in **Part 5: MPEG-7 Multimedia Description Schemes** of the MPEG-7 standard are organized in the areas shown in Figure 2-7 below:

- **Basic Elements:** this component includes the fundamental constructs which are used repeatedly throughout the definition of the MPEG-7 DSs. They provide specific data types and mathematical structures, such as, vectors, matrices, and histograms, which are important for audiovisual content description. Schema tools are intended to facilitate the creation and packaging of MPEG-7 descriptions.
- **Content Description:** the component that describes the structure (regions, video objects, frames, and audio segments) and semantics (objects, events, abstract notions) of audiovisual content.
- **Content Management:** this component handles various aspects of creation and production, media coding, storage and file formats and content usage.
- **Content Organization:** it is required for the organization and modeling of

collections of audiovisual content.

- **Navigation and access:** it facilitates browsing and retrieval of audiovisual content by defining summaries, partitions, decompositions, and variations of audiovisual material.
- **User Interaction:** it is used for the description of user preferences pertaining to the consumption of the multimedia material.

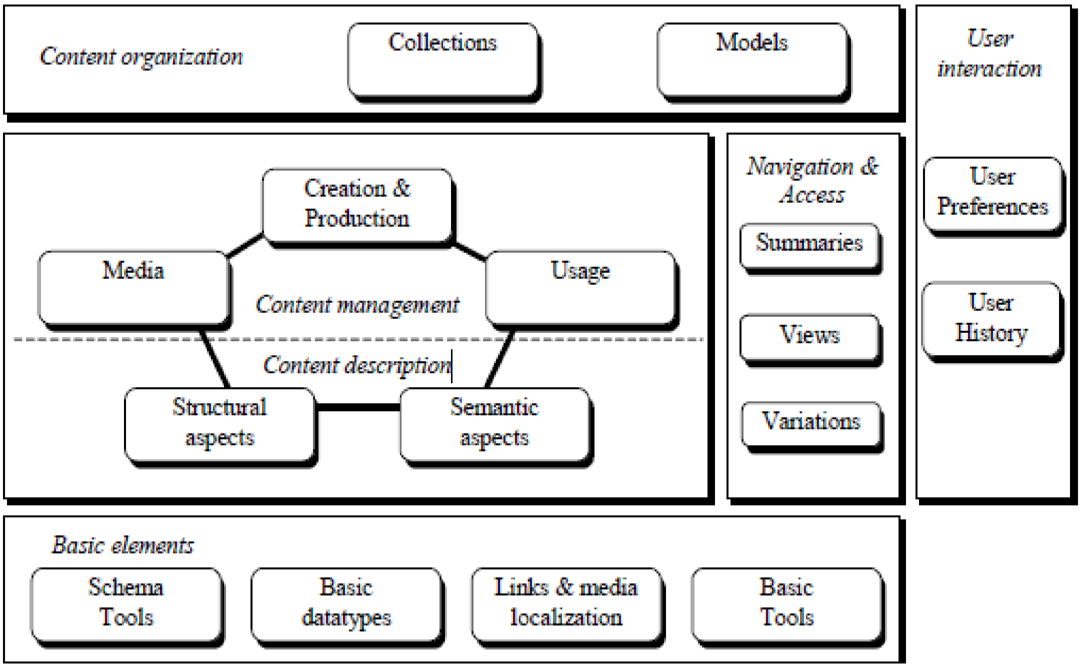


Figure 2-7: Overview of the MPEG-7 MDS Schemes (Source: (Chang, Sikora and Puri, 2001))

- **Part 6: MPEG-7 Reference Software:** includes the software implementation of the tools specified by the first five parts of the MPEG-7 standard.
- **Part 7: MPEG-7 Conformance Testing:** includes guidelines and procedures for testing the conformance of MPEG-7 implementations and streams.
- **Part 8: MPEG-7 Extraction and Use of Descriptions**: it provides information on the extraction and use of some of the description tools. Notably, it gives insight into the reference software and alternative approaches.

2.4 Related Work

The section below describes the different systems, tools, and frameworks we researched into during the course of this work. It details the MPEG-7 compatible systems in use to date and some of the non-MPEG-7 compatible tools that have been developed.

2.4.1 MPEG-7 Compatible Systems

There exists a variety of MPEG-7 enabled systems and applications to date. We just highlight on a few of the systems/tools worth mentioning below:

2.4.1.1 Frameline 47

Frameline 47 is a tool created by the Versatile Delivery Systems Limited. It is a MPEG-7 based system and has four main modes, which are File, Edit, Notate, and Present. The File mode allows a user to encode video files to Apple MPEG-4 H.264 format, convert MPEG-4 based .mov files to the .mp4 file format and allows a user to search for specific files within a given folder. The Edit mode basically allows users to edit a video, e.g. chop down a video into segments both manually or automatically. It allows for the creation of groups, and segments in videos. The Notate mode allows for the addition of descriptions, tags, and/or annotations to individual segments and/or groups. The present mode is focused on the presentation of the media (video, files, etc.) to the user. It allows for the viewing of videos in fullscreen mode, filtering of files according to their content, viewing the names of segments and groups during playback, etc. (Versatile Delivery Systems Ltd, 2006)

2.4.1.2 BilVideo-7

This tool is a MPEG-7 enabled, distributed, video indexing and retrieval system that supports complex multimodal queries in a unified framework (Bastan, Cam, Gudukbay and Ulusoy, 2010). BilVideo-7 gets its name from its predecessor BilVideo, which was a prototype video database system that supported the querying of videos using keyword-based spatiotemporal queries. It used a knowledge base and a Prolog inference engine. BilVideo-7 was created by assimilating a MPEG-7 profile for video representation that enables detailed queries on videos and used their MPEG-7 compatible video feature extraction and annotation tool to obtain MPEG-7 compatible video representations. It also boasts a visual query interface designed to formulate complex multimodal queries and supports several MPEG-7 descriptors. The BilVideo-7 system aims to address the two major problems that MPEG-7 compatible systems have, the use of a coarse image or video representation, extracting low-level descriptors from whole images or video frames and annotating them, but ignoring region-level descriptors (Bastan, 2010). The BilVideo-7 system also addresses the failure of most systems to perform complex multimodal queries by combining video segments and descriptors in different modalities. However, BilVideo-7 is a complex system and cannot be used by novice users who are not thrilled by the concept of even adhering to standards.

2.4.1.3 IBM MARVel

This is an image and retrieval tool that provides automatic indexing and categorizing of image and/or video collections based on integrated content analyzing techniques (Tochtermann et al., 2007). MARVel provides retrieval strategies based on feature descriptors or semantic concepts. It supports a rich set of compression formats. However, the tool does not provide means for any data transfer. Moreover, the annotation domain is limited to the available concepts e.g. indoor, and outdoor. No extensibility facilities have been documented on MARVel, making it difficult to customize. It runs on a windows platform and requires memory space on the local disk.

2.4.1.4 MPEG-7 Library

The MPEG-7 Library is a C++ implementation of the MPEG-7 standard ISO/IEC 15938:2001 (Joanneum Research Institute of Information Systems and Information Management, 2013) document. It is an open source tool developed by The Joanneum Research Institute at Graz in Austria. It supports both Windows and Unix operating systems. The library can be used for the manipulation, validation, creation, etc. of MPEG-7 descriptors. However, the library does not contain feature extractors. It, therefore, supports developers in XML-DOM programming of MPEG-7 documents. The MPEG-7 library is a comprehensive tool in the sense that it supports annotation of all features for all available media, be it video, audio, images, and almost all MPEG-7 Description Schemes and Descriptors. These Ds and DSs are around 1200 classes. It is no surprise that the implementation of all these classes is sophisticated; therefore it is an advantage to obtain a library such as this. The library is extensible, hence allows for the creation of more D and DS. This is the major reason why we decided to make use of the MPEG-7 Library in our work.

2.4.1.5 IBM's VideoAnnEx Annotation Tool

This tool enables users to annotate video sequences with MPEG-7 metadata. Each shot gets represented by a single keyframe and can be annotated with static scene descriptions, key object descriptions, event descriptions, and other custom lexicon sets that the user might provide. However, this tool is limited to concept annotation and cannot extract low-level MPEG-7 descriptors from the video.

2.4.1.6 VIZARD – Video Wizard

The VIZARD tool is a video publishing tool that targets novice users for processing their home videos. It allows for the annotation of videos by introducing a video-book model that provides a structuring in chapters, sections, index, conclusions, and so forth. It supports the annotation of videos of MPEG and AVI compression formats. The tool is available to Windows users and is not documented which is a great setback for such a tool. It makes use of semi-automatic annotation and does not allow for low-feature extraction. It allows for the creation of well-formed MPEG-7 descriptions which unfortunately are not valid due to a MediUri description which does not conform to the standard. VIZARD provides no interfaces but enables data transfer by MPEG-7 instance documents and Annotator XML files (Tochtermann et al., 2007).

2.4.1.7 MPEG-7 Audio Analyzer

This tool allows the extraction and annotation of all available MPEG-7 low-level audio descriptors. Unfortunately, only a web-based online version is available. It is restricted to WAVE AND MP3 compression formats and supports the automatic extraction of all low-level audio features (Döller and Lefin, 2007). Its input audio file is restricted in size to 1 MB for WAVE and 300KB for MP3. It has well-formed MPEG-7 descriptions that are also not valid like VIZARD. It also does not have documentation and as such drops to the bottom of the list when considering tools to use for media annotation. It has no interfaces but has no restrictions in regard to domain and extensibility.

2.4.1.8 VizIR

The VizIR framework was developed at the Technische Universitat Wien in Austria. The project provides an open extensible framework containing approaches for feature extraction, distance metrics, and components of user interfaces (Döller and Lefin, 2007). The main targets of the framework are the creation of a basis for content-based retrieval applications and for research on new approaches for automatic content extraction within images and videos. The VizIR framework sounds promising but falls short of the MPEG-7 Library as it currently caters for around 800 MPEG-7 Descriptors and Description Schemes in comparison to MPEG-7 Library's 1200 global variables.

2.4.1.9 The M-OntoMat-Annotizer software tool

This tool aims at linking low-level MPEG-7 visual descriptions to conventional Semantic Web ontologies and annotations (Döller and Lefin, 2007). The visual descriptors are expressed in Resource Description Framework (RDF Working Group, 2014).

2.4.2 Other Multimedia Content Description Tools

Apart from MPEG-7 based tools, there exists a number of tools for the annotation and description of multimedia content. Bargeron et.al. developed a Microsoft Research Annotation System (MRAS), which is a web-based system for annotating multimedia web content (Bargeron, Gupta, Grudin, and Sanocki, 1999). The annotations it can handle include comments and audio in the distance learning scenario. Steves et.al. developed a Synchronous Multimedia and Annotation Tool (SMAT). SMAT is used for the annotation of images (Steves, Ranganathan and Morse, 2001). It has a shortfall in the sense that, there is no granularity for video annotations nor controlled-term labels. The European Cultural Heritage Online (ECHO) developed a multimedia annotation tool which allows people to work collaboratively on a resource and add comments to it. However, all these systems and tools are not MPEG-7 compatible. All is not in vain though as these tools give us a sense of direction concerning the capabilities we would like TISPAN standardized tools to have.

2.5 Summary

This chapter gave a detailed review of literature and theoretical background on metadata creation and the tools associated with its creation. The current state-of-the-art networks and tools were highlighted with major emphasis given on the IPTV architectures and the IPTV platform in general. The MPEG-7 standard was presented and its various components described in-depth. The chapter went on to discuss the various MPEG-7 compatible tools in existence to date, the non-MPEG-7 compliant tools and highlighted their benefits and shortfalls. The development of a tool that can implement the MPEG-7 standard for the description of multimedia content without the user being fluent with MPEG-7 standard technicalities would be a great contribution to the IPTV industry. Therefore, the development of the MCDMS prototype in this research is a worthwhile exploit. The following chapter outlines the research methodology and research design employed for the duration of this research.

3. Chapter 3: Research and System development methodology

This chapter gives the methodology framework followed in the design of the Multimedia Content Description and Retrieval Tool for IPTV. Presented, are the software development processes and tools employed in this work. The chapter introduces methodology, then goes on to explain the research design before further elaboration on the methodology is given. The chapter highlights the steps taken in undertaking this research and concludes with an overview of the chapter itself.

3.1 Methodology

Avison and Fitzgerald in their book (Avison and Fitzgerald, 2006) define a systems development methodology as *"a means to achieve the development, or part of the development, of information systems based on a set of rationales and an underlying philosophy that supports, justifies and makes coherent such a recommendation for a particular context. The recommended means usually includes the identification of phases, procedures, tasks, rules, techniques, guidelines, documentation, and tools. They might also include the recommendations concerning the management and organization of the approach and the identification and training of participants"*.

Kothari in his work (Kothari, 2004) stipulates that there exist two approaches to research, namely, quantitative and qualitative. He goes on to state that the quantitative research approach can be further broken down to include inferential, simulation and experimental approaches to research. The inferential approach entails the formulation of a database from which the characteristics or relationships of a sample population are deduced. It is a common procedure to infer the characteristics through observations or questionnaires. The simulation approach involves the construction of an artificial environment that is meant to handle complex phenomena and usually cannot be setup in a laboratory environment. This allows for the observation of the dynamic nature of systems and their subsystems in a controlled environment. The experimental approach involves the manipulation of certain variables in a much-controlled environment to observe what effect the variables have on other variables. It is safe to say experimental research involves the use of empirical tools to gather results from a real-world testbed. Experiments are carried out to qualify the truthfulness of an adopted hypothesis. The experimental research approach is one of the most commonly used research approaches in computer science for functionality, performance, and behavior analysis. In this study, we adopted the experimental research method. Figure 3-1 illustrates the experimental design process implemented during this study.

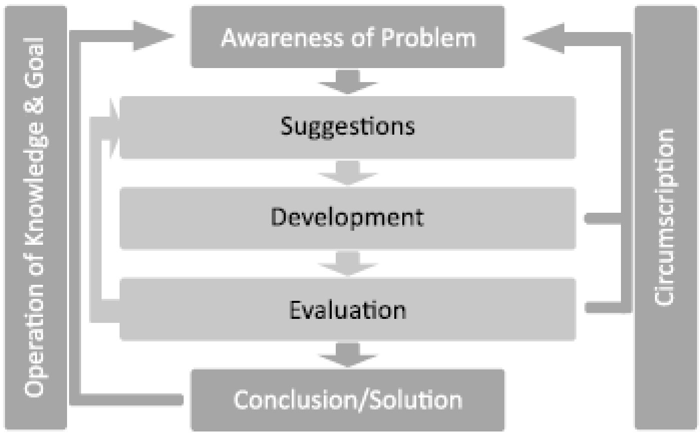


Figure 3-1: Research Design

The design process as Takeda et. al. analyzed brings forth the idea that an awareness of a problem should be at the forefront of the research model (Takeda, Veerkamp, Tomiyama and Yoshikawa, 1990). Suggestions to solve the problem are then drawn up logically. It is these very suggestions that mesh up to form a hypothesis. When hypotheses are drawn, an artifact is developed that will aim to implement the proposed solution. Implemented solutions are then evaluated. The Suggestion, Development, and Evaluation phases are often performed in an iterative manner in an attempt to produce more accurate results. Termination of the project occurs when conclusions are reached. At the end of each project, we would have gained new knowledge. The new knowledge production is shown in Figure 3-1 as Circumscription and Operation of Knowledge and Goal. In design research, "circumscription" has been noted as a relevant process as it outlines the importance of construction to gain understanding.

3.2 Research Design

The research methodology adopted in this work has been discussed in section 3.1. This subsection provides insight into the tools and techniques employed when breaking down the research design as described above. Sections 3.2.1 to 3.2.5 describe the research design that we implemented as illustrated in Figure 3-1.

3.2.1 Awareness of Problem

Our focus was on IPTV technology. It was recognized that IPTV is an emerging technology with the capability to be the next-big-thing (JSB Market Research, 2014). However, IPTV brought with it complications concerning how to access a particular multimedia content uploaded. We needed to know how content is saved on a database before we are able to retrieve that content. Therefore we needed to know how content is described and saved on IPTV servers. Therefore, we undertook a literature review on multimedia content description. We researched into the advancements that have been taken in attempting to offer the best IPTV services to end-users. As highlighted in section 2.1, SDO's have combined efforts to ensure that state-of-the-art technologies are developed with consideration given to standards. We highlighted that NGN's have been created by various SDO's and went on to stress how much the introduction of IMS has influenced the creation of IPTV services on the fly. Nowadays, users can easily access tools that can be used to create multimedia content. This content is what we have referred to as User Generated Content (UGC) in the preceding chapters. The exponential growth in UGC uploads is a major concern, there now exists an abundant supply of multimedia content that is difficult to retrieve when searched for because the content is poorly described and annotated. Therefore, we stressed the need for standardized multimedia content description to aid in the retrieval of such user-generated multimedia content. The researcher took an online course offered by the University of North Carolina at Chapel Hill entitled Metadata: Organizing and Discovering Information. In this course, various content description tools and standards were studied. Content Description standards such as Dublin Core, Text Encoding Initiative, Metadata Encoding and Transmission Standard, and the Multimedia Content Description Interface were studied, just to mention a few. Project Isizwe: The Tshwane Free WiFi project encouraged the dissemination of free WiFi to the residents of South Africa. Project Isizwe justified our concerns over the standardization of multimedia content description. Recently, the Trump archive was launched (Kaplan, 2016). The archive contains over 700 speeches, interviews, and debates that the president of the United States of America gave while he was campaigning for the presidency. The content is more than 520 hours long and all this information can be manipulated by various users to create millions of video bits that without proper description could add to the already infinite multimedia content on the www that is difficult to retrieve. This is just one example, imagine how much more multimedia content is on the Internet that is cumbersome to retrieve when you consider how many more influential, radical, and controversial figures this planet possesses.

3.2.2 Suggestions

In a bid to solve the problem we had to go through relevant literature, IPTV standard documentation was the first port of call as it contains prescribed solutions. TISPAN documentation settled the debate concerning which multimedia content description standard we should adopt. The Multimedia Content Description Interface (MPEG-7) is the standard that has been adopted by TISPAN for the description of multimedia content. The MPEG-7 standard has been described in detail in section 2.3 of this thesis. We believed a MPEG-7 enabled tool that adhered to the TISPAN specification documents would go a long way in ensuring that the retrieval of multimedia content would be much more efficient if the tool was used for the standardized description of the created multimedia content. According to the MPEG-7 standard, MPEG-7 enabled applications can encode and decode MPEG-7 descriptions which can be stored together with the multimedia content or separately, thereby influencing the search and retrieval of such multimedia content. Therefore, it was logical to assume a MPEG-7 based Multimedia Content Description and Retrieval tool would solve the multimedia content retrieval problem caused by the lack of standardized multimedia content in archives, databases, and the Internet. It was, therefore, our hypothesis that the MPEG-7 based Multimedia Content Description and Retrieval Tool would enable the seamless search and retrieval of multimedia content (IPTV services) offered by the IPTV platform.

3.2.3 Development

This phase of the research employed a combination of two different types of software engineering, namely, incremental development and reuse-oriented software engineering. The former is mostly based on the notion of developing an initial implementation or prototype and exposing it to the users, then evolving the prototype through several versions until an adequate system has been developed. However, we do not involve any users in this particular work; the initial implementations are vetted by the researchers and altered accordingly. Reuse-oriented approaches rely on a majority of reusable software components and an integrating framework for the composition of these components. At times these components are systems in their own right that may provide specific functionality (Sommerville, 2011).

3.2.3.1 Incremental Development

The incremental development technique is a combination of both linear and iterative types of development. Linear development involves the breakdown of a project into sequential phases, whereas, iterative development entails the repetition of certain phases of the development to produce a better product using knowledge acquired from the previous build. The main goal of incremental system development is to reduce inherent project risk by breaking a project into smaller segments and providing more ease-of-change during the development phase. The initial software concept, requirements, and design of the architecture of the system core were defined using the Waterfall approach, followed by iterative prototyping, which resulted in the final prototype (annotation tool). Incremental development has long been hailed as an ideal technique for handling the development of Web-based Information Systems and event-driven systems (Centers for Medicare & Medicaid Services, 2008). To that end, we found it as an adequate software development technique for this particular study.

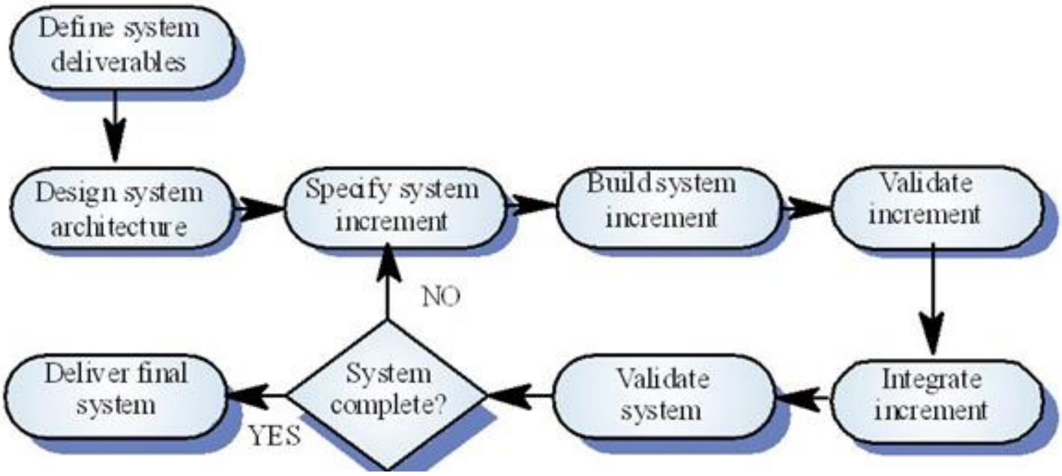


Figure 3-2: Incremental development process (Source: (Sommerville, 2000))

3.2.3.2 Re-use oriented Development (ROD)

Re-use oriented development is a method of software development where a refined system is developed from a series of prototypes that preceded the refined system. ROD follows a sequence of defined rules in its attempt to produce a refined and complete system/tool. Ian Sommerville in his book (Sommerville, 2011) points out that at times the software components involved in ROD are systems in their own right that may provide specific functionality. As depicted in Figure 3-3, the first stage of ROD is to gather system requirements. Secondly, the discovery of software that could handle the requirements specified requirements. This software is evaluated with regards to the requirements. Upon discovery of the specific software in existence, we then refine our requirements and iterate if necessary. If a system exists that can do as required we configure the system and evaluate. If a system exists that has part of the functionality required we may adapt some of the system components and develop some other components to amalgamate with the existing software. When the components are completely developed and have full functionality we then integrate the system components and evaluate the system against the specified requirements. In this research, the software that we discovered and used as system components are the MPEG-7 Library, FFmpeg (FFprobe), and a native XML database. Figure 3-3 below gives the flowchart of ROD.

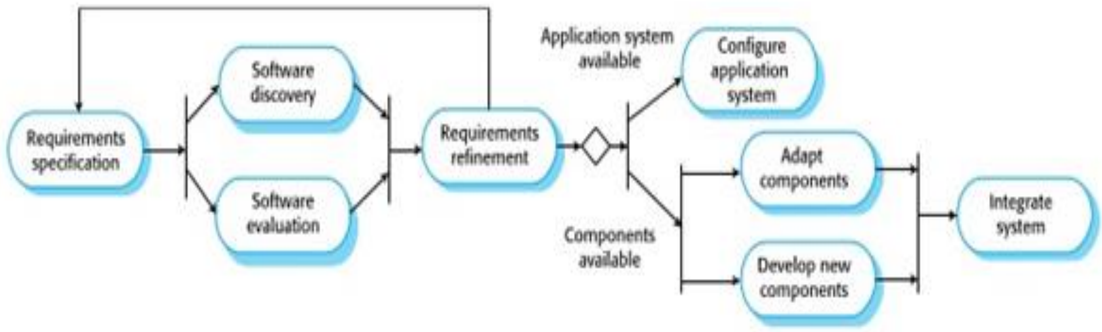


Figure 3-3: Re-use oriented development model (Source: (Sommerville, 2015))

3.2.3.3 The Development tools

A variety of tools were used for the development of our tool.

- Xerces 3.1.1
- Cmake 3.5.0
- MPEG-7 Library
- FFmpeg (FFprobe) 3.0.1
- Native XML Database (Sedna)

A detailed description of the above-mentioned development tools is given in Chapter 4 section 4.2.

3.2.4 Evaluation

To prove whether the developed tool could meet the requirements specified we undertook a series of tests. The system had to prove it could extract technical information about a particular content that an amateur content producer would not know e.g. bit stream filters, codecs, pixel formats. The GUI associated with the Description Module should allow the user to provide a semantic description of the content. All this information should then be encoded into the MPEG-7 format and stored in a native XML database that an IPTV system would understand. The MPEG-7 descriptions created and stored in the native XML database should be retrievable when another MPEG-7 enabled application requests the multimedia content.

3.2.5 Conclusion/Solution

At this phase of the research, we deduced that our prototype (tool) had solved our stated problem and we could, therefore, state that our hypothesis had been proven accurate as our tool could elaborately annotate multimedia content with metadata that could be retrieved efficiently when searched for. The system allows for the annotation of multimedia content thereby creating MPEG-7 descriptions of multimedia content that can be stored together or separate from the multimedia content. These MPEG-7 descriptions can be searched for and retrieved easily; hence the requirements of IPTV search are fulfilled.

3.3 Summary

This chapter focused on the research and systems development methodology. The chapter underlined the research design adopted for this research and went on to discuss the distinct components of the research design from section 3.2.1 to section 3.2.5. The systems development methodology employed was stated and elaborated upon. Chapter 4 presents the design and implementation of the MCDMS.

4. Chapter 4: Design and Implementation

In this chapter we present an overview of our approach to the design and implementation of the MCDMS highlighted in Figure 2-4 of Chapter 2. We begin by stating the different system components that were used during the development of the prototype. A clarification on the system architecture is presented. Thereafter, we elaborate on the different system components and how they fit together to form a refined tool that can be used for the description of multimedia content suitable for IPTV service provisioning. A clear picture of how the MCDMS is meant to work is painted before we set off to describe how it was implemented. The different platforms, frameworks and their requirements are detailed in this chapter as we aim to explain how to setup this system when in possession of the different system components as required.

4.1 Design

As mentioned in section 3.2.3, we employed a combination of incremental development and ROD during the entirety of this research. We avoided the need to re-invent the wheel and, as such, managed to save up on time as some of the software required is quite complex and sophisticated. We made use of a variety of open source tools to implement the functionality requirements of the system components where necessary.

4.1.1 System Architecture

A system has been defined by the IEEE Standard 1472000 as a collection of components that have been structured to realize a specific function or set of functions (IEEE-SA Standards Board, 2000). The standard goes on to define architecture as *the fundamental organization of a system embodies in its* *components, their relationships to each other, and to the environment, and the principles guiding its design and evolution*. Therefore, the system architecture design process established a guiding structural framework for the MCDMS. The general system architecture is shown in Figure 4-1 with its various components or subsystems that we discuss in section 4.1.2.

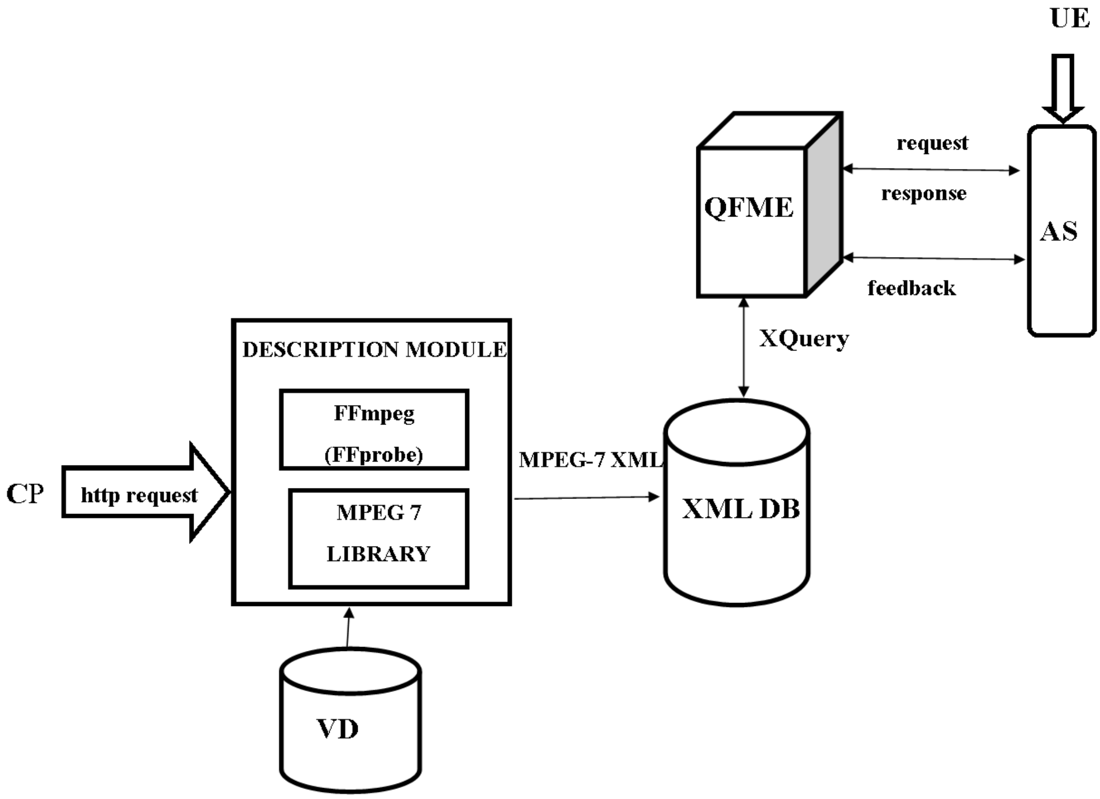


Figure 4-1: Overview of the System Design (Source: (Ncube and Shibeshi, 2015))

4.1.2 System Components

As can be construed from the system architecture, a variety of tools and systems were used to develop the MCDMS. The MCDMS is distributed client-server architecture. The client side consists of the Content Provider, and the users making use of the User Equipment. The server side comprises of the Description Module, the Query and Feedback Management Engine, the Sedna Native XML Database, and the Video Database. The following sections discuss these components in brief.

4.1.2.1 Content Provider (CP)

Nowadays, the CP can be anybody who possesses a technological device that can produce any multimedia content e.g. a mobile phone, a digital printer, a scanner, etc. Technology has advanced to the extent that not only big broadcasting companies produce content that is consumed by the general public, but the general public is also producing multimedia content. We now have terms such as “prosumers” that have been coined to describe the consumers of today’s content. These consumers have turned out to be producers of content as well hence the term “prosumers”; they are both producers and consumers of multimedia content. The IPTV specifications cater for the provisioning of User Generated Content (UGC) hence prosumers are here to stay. The CP can request the MCDMS for a video description. The system through the Description Module provides an interface to the CP that allows the CP to describe the video they want to upload, thereby providing the semantic information about the uploaded video for easy access by users.

4.1.2.2 Description Module (DM)

The DM is a composite of the MPEG-7 Library and FFmpeg. One of the goals of this research was to be able to create MPEG-7 descriptions of multimedia content (in this case video) hence the adoption of the MPEG-7 Library. The library is a MPEG-7 based feature extraction and annotation tool. It has the ability to encode and decode multimedia content descriptions into the MPEG-7 format making use of the Detailed Audio-Visual Profile (DAVP) (Joanneum Research Institute of Information Systems and Information Management, 2009). The MPEG-7 Library has the ability to manipulate these multimedia content descriptions, serialize them to XML and de-serialize them with validation from XML using Xerces (The Apache Software Foundation, 2016). The library guarantees the use of standardized multimedia content descriptions as it facilitates such descriptions by providing over 1200 Descriptors and Description Schemes as specified by the MPEG-7 standard. Therefore, MPEG-7 Library is used for analysis of the uploaded multimedia content, annotating the content and then serializing/encoding the description that has been created into the MPEG-7 format. FFmpeg has been adopted for its FFprobe component. FFprobe has the ability to extract the technical information associated with a video such as codecs, bit stream filters, size, pixel formats, etc. and the CP need not know the technical information. FFprobe extracts the information and displays it for the user to see, the information is appended to the description offered by the CP and formatted accordingly by the MPEG-7 Library. Therefore, the DM succeeds in the semi-automatic description of multimedia content (video) and the encoding of such information into the MPEG-7 format which happens to be XML format since the MPEG-7 standard adopted XML as the primary schema.

4.1.2.3 Video Database (VD)

There exist a variety of video collection databases we could use, e.g. Tellico, Griffith, vMovieDB, etc. however, in this work we made use of a particular directory within our Personal Computer.

4.1.2.4 Extensible Markup Language Database (XML DB)

The XML DB that we adopted for this research was the Sedna Native XML Database (Institute for System Programming RAS, 2012). Sedna is a free native XML database which provides a full range of core database services. We chose Sedna because just like the MPEG-7 Library it is implemented in C/C++, it has support for W3C XQuery language validated by W3C XQuery Test Suite; it has tight integration of XQuery with full-text search facilities and a node-level update language. The XML DB stores the MPEG-7 XML-based metadata about the video content being uploaded onto the IPTV platform. The MPEG-7 metadata, which is stored in a hierarchical XML representation can be manipulated in an object oriented hierarchical class tree which can be modified, or deleted, and new elements can be created. Therefore, the XML DB will contain various XML files with regard to differing metadata elements, example a presence element, an action element, and a reference link element to the video itself, to name just but a few.

4.1.2.5 Query and Feedback Management Engine (QFME)

The QFME makes use of XPath, the XML Path Language which is a query language for selecting nodes from an XML document. It makes particular use of the XQuery function to request and retrieve particular XML document-based metadata. The QFME is the intermediary between the AS and the XML DB. It is responsible for the management of queries and the feedback mechanism of metadata. It can also serve for load balancing purpose.

4.1.2.6 Application Server (AS)

Metadata-related requests come from the AS. XML-based queries are brought forth to the QFME via the AS by the users of IPTV. It should also be noted that the AS itself could also ask for metadata. The requests and responses are done through signaling protocols specific to IPTV. The protocols include the Real Time Streaming Protocol (RTSP), and the Session Initiation Protocol (SIP). Both push and pull technologies are implemented in this design which allows for feedback to be given to the MCDMS thus updating metadata about each video XML file in the XML DB.

4.1.2.7 User Equipment (UE)

The UE provides a consumer with access to the IPTV platform via the AS. It can refer to the set-top box, the remote control or even the television used to view the output from the IPTV platform. We are not entirely concerned about the type of UE that a user might use to access the IPTV platform. We focus mostly on how the user describes his/her uploaded video content and how they search and recover their content.

4.2 Implementation

Section 4.1 dwelt on the tools that were used to build the prototype MCDMS. This section is devoted to describing how these tools were used in the implementation of the system. The system prototype that we describe in this section serves as a proof of concept and a solution to the problem highlighted in Chapter 1 section 1.2. Each of these tools has their own requirements, i.e. specific requirements for the development environment. The specific requirements are highlighted in the sections that follow.

4.2.1 Setting up Xerces

Xerces is a validating XML parser written in a portable subset of C++. Xerces-C++ makes it easy to enable applications with the ability to read and write XML data. Xerces is used by the MPEG-7 Library in this research for serializing and de-serializing multimedia content descriptions. Xerces can be installed via the Synaptic Package Manager or from source using the Linux terminal. It is best to install Xerces from the terminal as it allows you to install the exact version required by your corresponding software (e.g. MPEG-7 Library requires Xerces v3.1.1). The MPEG-7 Library samples we use during this research require the Xerces root directory to be on the same level as the MPEG-7 library root directory.

4.2.2 Setting up Cmake

Cmake is a meta-build tool that is used to generate native build systems. It should be noted, however, that it is not an actual build tool but it is used to generate MakeFiles for Visual Studio 6 and upwards. We setup Cmake version 3.5.0 for the generation of MakeFiles that were used by the MPEG-7 Library framework. The Synaptic Package Manager was used for the setup.

4.2.3 Configuring the MPEG-7 Library

The MPEG-7 Library can be found in source or in binary format. The requirements of the development environment are:

- GCC version 4.1.0 or higher
- Xerces version 3.1.1
- The sample project we use assumes that the Xerces is installed at usr/local/include/xercessc and usr/local/lib

We describe how we configured the MPEG-7 Library from source. The library comes with a manual, however, it is quite complicated for novice users to know how to apply the instructions hence we point out how to configure the library step by step:

- Open the terminal
- Change directory to wherever you saved the MPEG-7 Library source file e.g. cd /home/prince/Desktop
- Create a debug and release directory
- Change directory to the debug directory you just created
- cmake –D CMAKE_BUILD_TYPE = Debug /home/prince/Desktop/Mp7Jrs2.7_src_win_mac_linux
- change directory to the release directory you created
- cmake CMAKE_BUILD_TYPE = Debug /home/prince/Desktop/Mp7Jrs2.7_src_win_mac_linux

- make
- make install

4.2.4 Installing FFmpeg

FFmpeg is a project/tool suite that comprises three main executables, FFmpeg for conversion, FFprobe for interrogation, and FFserver for streaming. For the purposes of this research, we needed to employ FFprobe for interrogation of the videos to be uploaded onto the IPTV platform. FFmpeg can be installed using the command: **sudo apt-get install ffmpeg**.

4.2.5 Setting up Sedna

Sedna is a Native XML database that is flexible, scalable, and interoperable. It has no capacity limits for data types and allows for structural and semantic searches. It allows for three types of queries, namely query, update, and bulk. These queries enable us to meet the functional requirements of the MCDMS developed in this research. We need to connect to the Sedna database:

```
struct SednaConnection conn = SEDNA_CONNECTION_INITIALIZER;


int SEconnect(SednaConnection* conn,
        const char* url,
        const char* db_name,
        const char* login,
        const char* password);


int SEclose(SednaConnection* conn);
```

4.2.6 The Graphical User Interface

When accessing the Multimedia Content Description and Management System (MCDMS) using the GUI associated with both the CP end and the QFME end we get the GUI illustrated below in Figure 4-2. The GUI was developed using the Netbeans IDE version 8.0. When the MCDMS Welcome Tab appears, the user then decides whether they want to upload a video or search for one. If the user is uploading a video they click on the “Upload Video” button which will then direct them to the MCDMS Video Upload Tab as illustrated in Figure 4-3. Thereafter, the user is required to completely describe the video they would like to upload onto the IPTV platform. A total of sixteen metadata elements have been listed on the MCDMS Upload Video Form, these are the sixteen most basic elements required when describing multimedia content. Some of the elements listed on the form may be unknown to the user hence they get access to the technical details of the video they would like to describe via FFprobe which returns the details of a video as requested by the user. In this work, we have a text file called “VideoInterrogation” which contains the technical details of a video file that has been requested. This text file will be described in more detail in Chapter 5 section 5.1. The user also has an additional text area where they can add more description about the video if they have more descriptions to add. At the bottom of the GUI, there is a button which you click in order to select your video from wherever you have it stored, to add it to the video database. Once the description has been completed, and the video selected, the consumer clicks on the submit button and the MPEG-7 Library part of the DM is brought into action. The annotated description is converted into an XML MPEG-7 format which will then be stored in the Sedna native XML database. A popup dialog box appears and confirms that the MPEG-7 description of the video has been uploaded and saved on the MCDMS. The exit button closes the MCDMS Welcome tab.

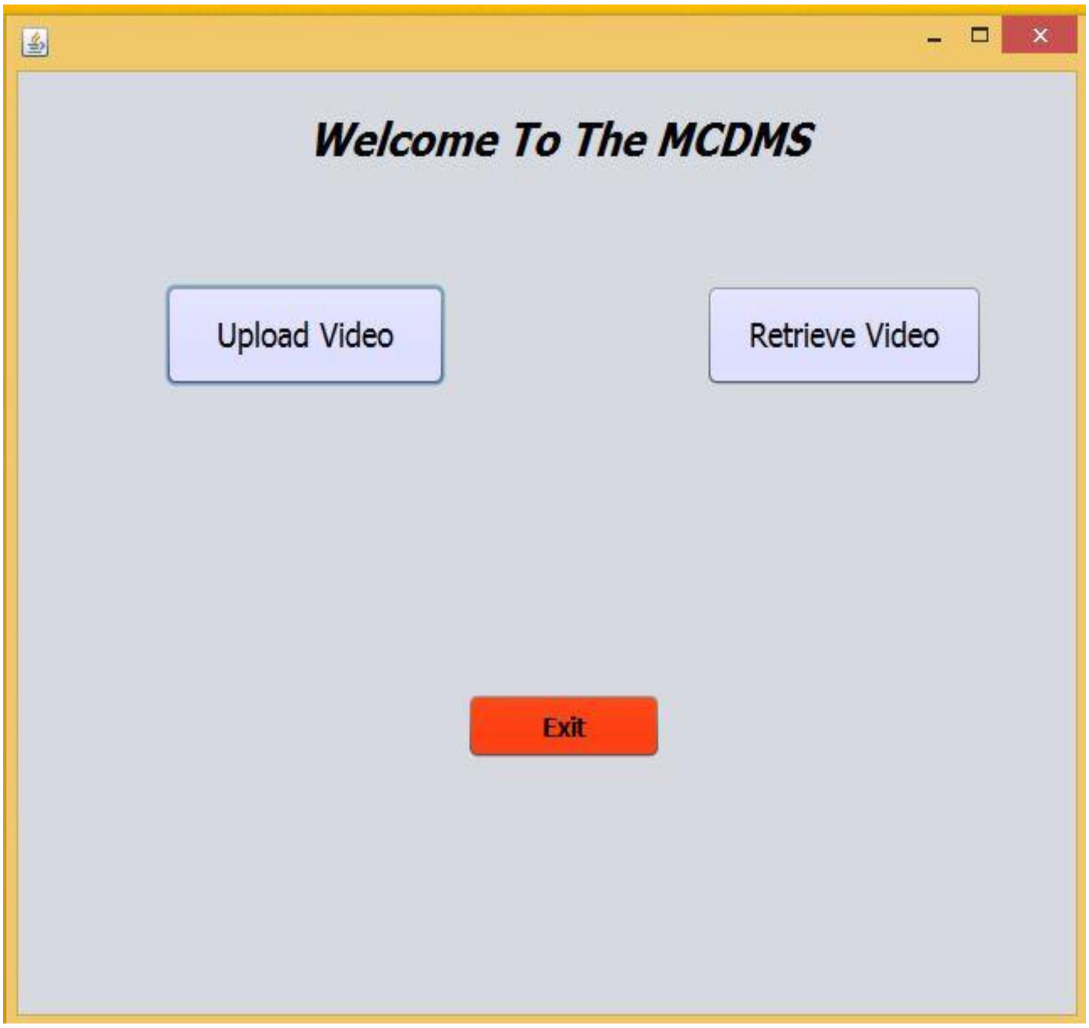


Figure 4-2: MCDMS Welcome Form

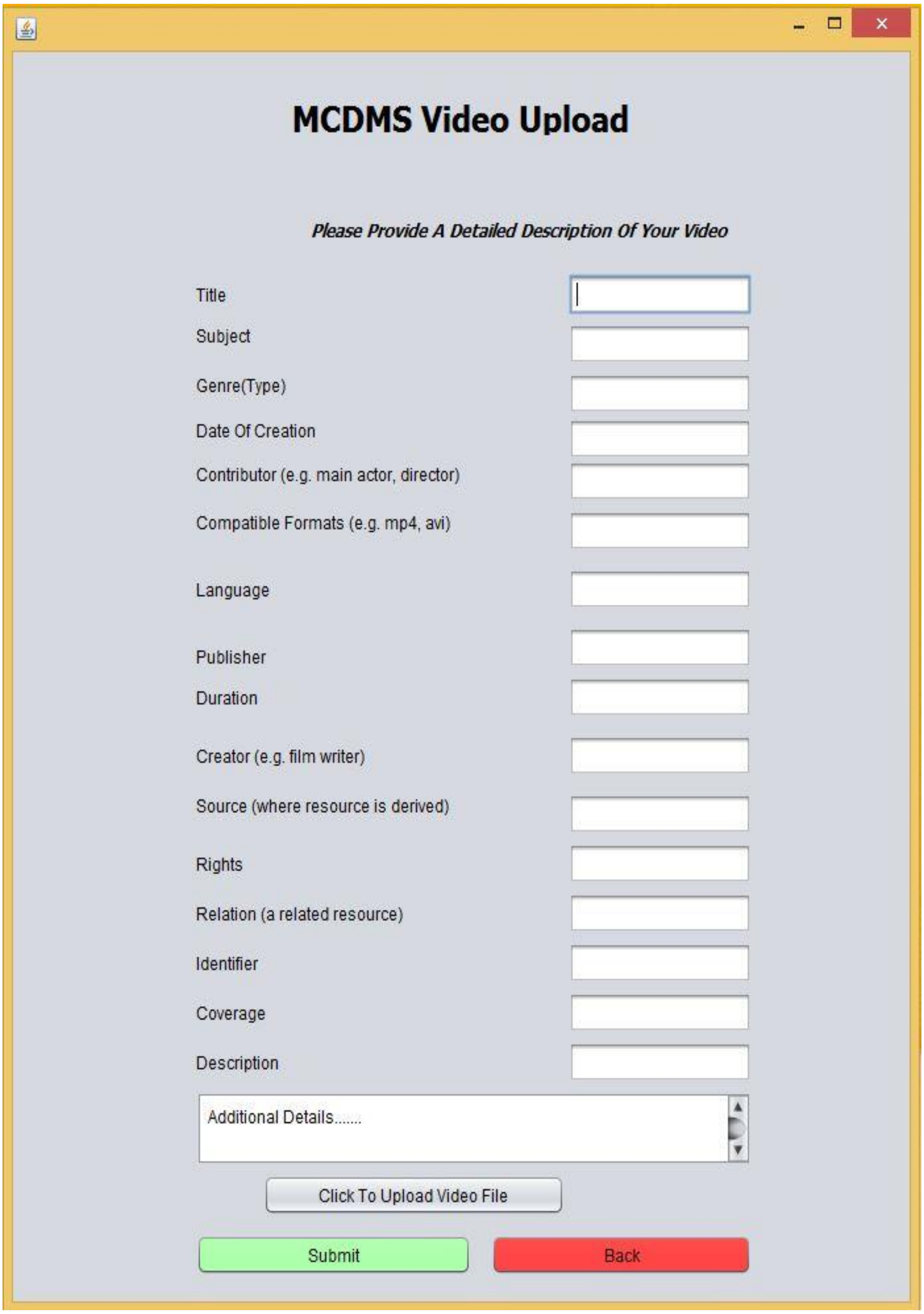


Figure 4-3: MCDMS Video Upload Form

The user may decide to search for a video from the video database. The MCDMS Welcome tab is accessed, and then the "Retrieve Video" button clicked. When the button is clicked, the MCDMS Video Search Form pops up. The form contains all the required sixteen basic metadata elements. This allows for the user to provide a comprehensive description of the video for the search of the required video to be as efficient as possible. The more the description offered, the better the search results yielded. After a description has been given, the user clicks the submit button and the QFME goes to work. The description is also converted into XML MPEG-7 format and a query logged via XQuery which then interrogates the components of the Sedna XML database and retrieves all the videos that are directly linked to the given description. The videos are retrieved and listed by title. The user needs only select the video they require from the returned list. The "Back" button returns the user to the MCDMS Welcome Tab.

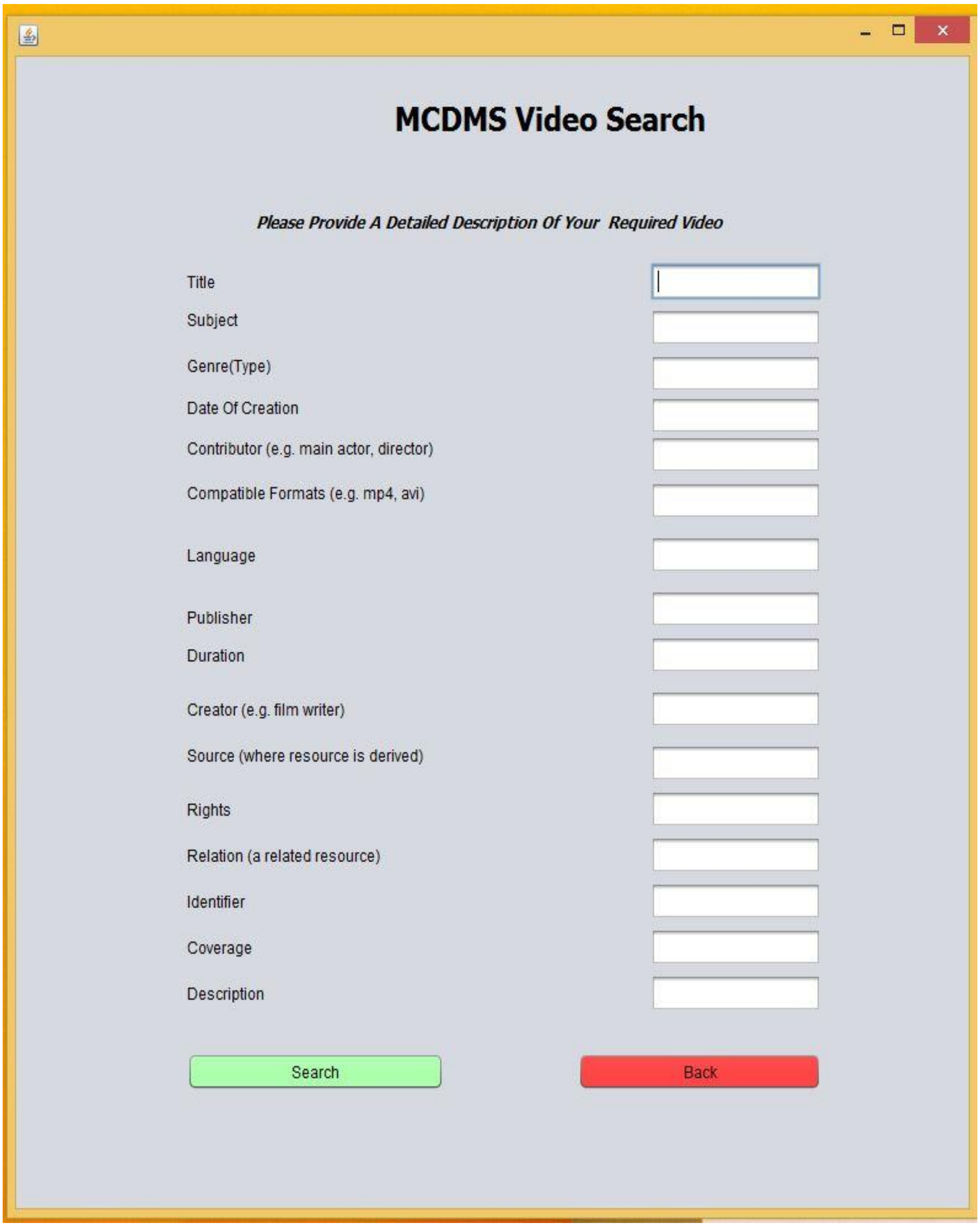


Figure 4-4: MCDMS Video Search Form

4.3 Summary

This Chapter has offered a detailed discussion of the system design, which includes a discussion about the overview of the system architecture. It went on to discuss the implementation stage of the MCDMS and concluded with a summary of the chapter. The ensuing chapter presents testing, results, and a discussion of the results.

5. Chapter 5: Testing, Results, and Discussion

This chapter presents and discusses the tests undertaken and the results obtained in this study. The chapter is sectioned as follows, section 5.1 presents the tests carried out, section 5.2 provides the results of the various tests carried out, section 5.3 interprets and analyzes those results, and section 5.4 summarizes the chapter.

5.1 Functional Testing

Functional testing has long been assimilated into software development as a software testing process whereby software is tested to ensure that it conforms with all specified functional requirements (Techopedia, 2017). Primarily, functional testing is used to verify that particular software produces the same output as required by the end-user. In an attempt to achieve such results, software is tested by providing it with some related input so that we can evaluate the output obtained in relation to the required base functional requirements. Functional testing has a number of techniques, including, smoke testing, black box testing, white box testing, and user acceptance testing. For the duration of this research, we had to undertake a series of smoke testing where we had to verify that the features of the software we used work properly. We undertook component testing whereby we tested individual units/components of the MCDMS to validate that the units performed as designed. The separate units we tested were the DM, the GUI, and the Sedna database. The DM was broken down into its two components, namely the MPEG-7 Library unit, and the FFprobe unit. The units were tested accordingly to validate whether they achieved the functionality they were designed to possess. We also tested the functionality of the system as a whole (integrated components working together) to ensure we had a fully functional MCDMS prototype.

5.1.1 Testing Individual Components

We tested both components of the Description Module. Firstly, we tested the FFprobe component by making use of various videos we had stored on the Acer PC running on a Linux 12.04 LTS OS that we used throughout this research. We ran the FFprobe interrogation command that returned all the technical information about the videos checked. We ensured the returned output would be saved to a text file named "VideoInterrogation". The text file is cleared every time a probe is run and the resulting information appended to the clear file.

We present two screenshots below in Figure 5-1 that highlight the results obtained from such probes. The "VideoInterrogation" text file contains all the technical information about the interrogated video that a user might not know. This information is then available to the user and can be copied onto the MCDMS Video Description Form.

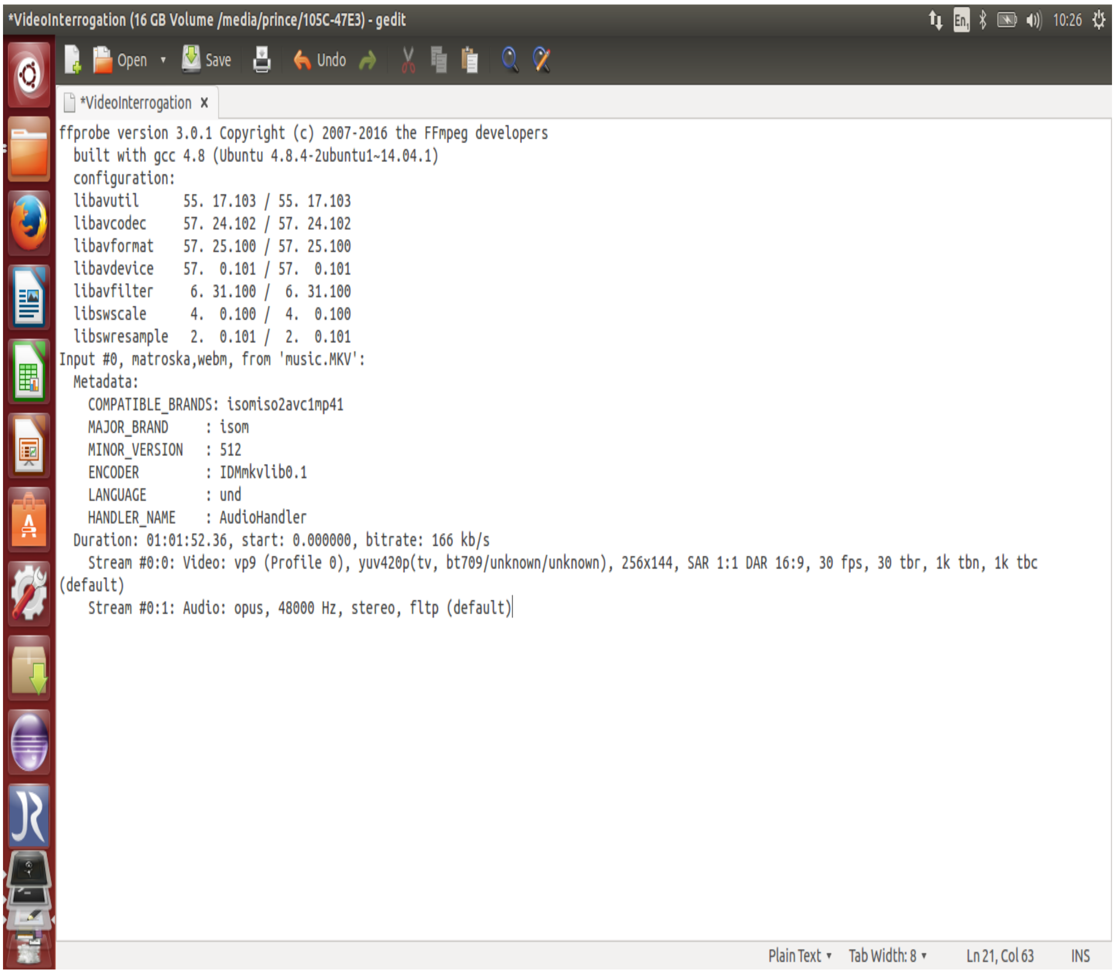


Figure 5-1: FFprobe Interrogation of a music video

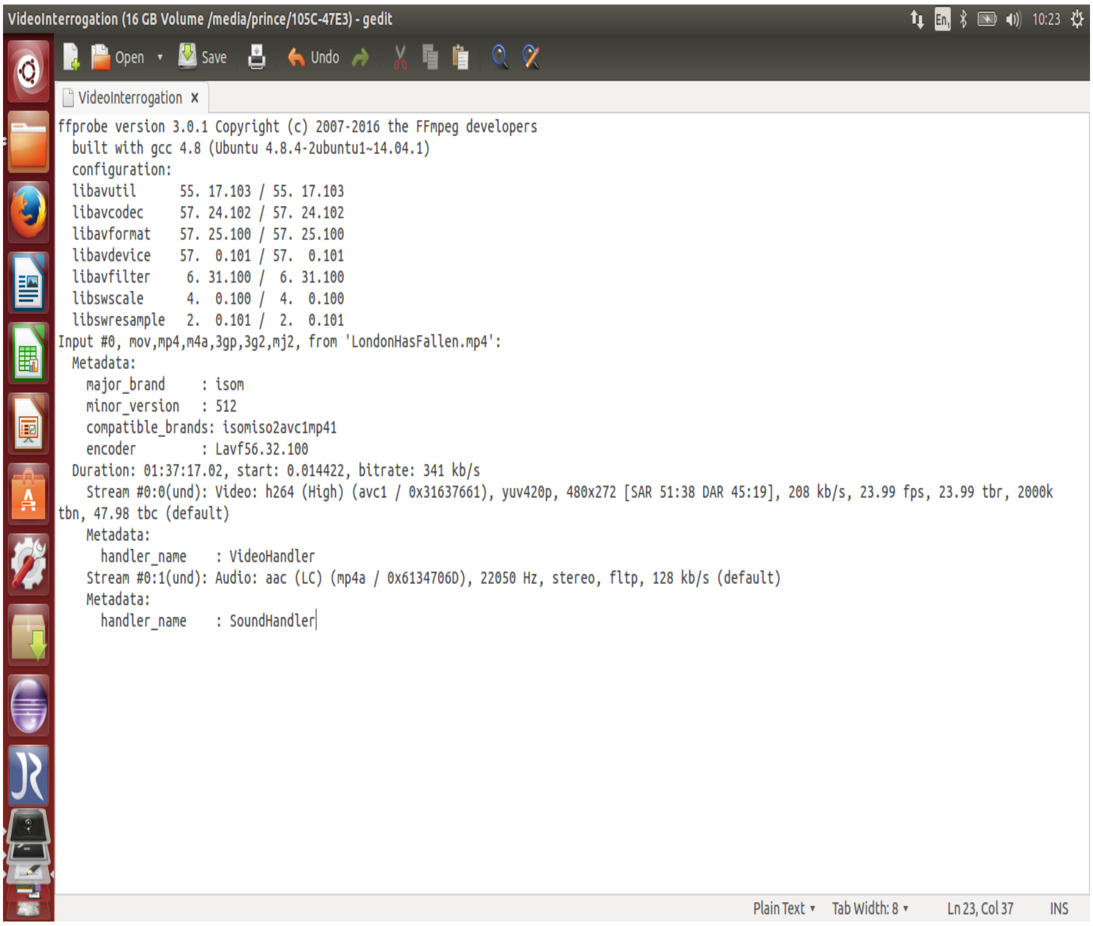


Figure 5-2: FFprobe Interrogation of a movie entitled London Has Fallen

The MPEG-7 Library functionality was tested using the sample programs that the library distribution comes with. We made use of the Sample programs entitled "Simple", "CreateDocument", and "ValidateDocument" to ensure that the component was functioning appropriately. The MPEG-7 Library formats the descriptions given to it into an XML MPEG-7 format by making use of the DAVP. Firstly, we created a new MPEG-7 description using the "Simple" program. This new program could be validated using the "Validate Document" program to prove that the document was indeed a MPEG-7 standard document. The programs create test objects that are XML MPEG-7 format documents that can be saved within the Sedna database.

The GUI was tested by uploading new videos onto the MCDMS system. One of the uploaded videos was an episode of the series entitled “House of Lies”. We gave the MCDMS a detailed description of the episode, and the description was accepted and saved by the MCDMS. We also searched for the video by giving an incomplete description of the video and then by giving a complete and elaborate description of the video. We iterated using different videos and tested the efficiency of retrieval when given a detailed description of the required video and when given a short description of the required video.

The Sedna native XML database was tested in terms of connectivity and querying.

5.1.2 Testing the Entire System (Integration Testing)

With the individual components tested and found to be bug-free, we moved on to integration testing. Integration testing is done to uncover defects in the interfaces and in the interaction between integrated components or systems. During the course of this research, we did not employ any external testers or evaluators to test the MCDMS or its components. We undertook our own testing and we used the Big Bang approach to integration testing. The Big bang approach entails the testing of all system components or most of the components after they have been combined. We tested the interactions between the system components to validate whether there was seamless communication between the various units. The DM was tested with regards to video annotation and encoding of video description to MPEG-7 format before being passed on to the Sedna database. The Sedna database was tested for its ability to receive and store new entries, update current entries, delete entries, retrieve entries when they are requested by an end user. The user interface associated with the CP and the QFME was tested to see whether it would launch the creation of video content metadata creation in MPEG-7 format by the DM when the MCDMS Video Description Form submit button is clicked on. The DM should kick into action as soon as the submit button is clicked on. The QFME makes use of the same user interface when a user needs to search for a video. The GUI was tested by entering a search for a comedy video by Mr. Bean entitled “Car_Park_Chaos”. Different descriptions of the video were given and results obtained for the queries. A total of 33 videos were uploaded onto the MCDMS. 15 of the videos were annotated in detail using the 16 basic metadata elements provided within the GUI, 12 were partly annotated with a maximum of 10 metadata elements each, and 6 were not annotated at all.

Table 5-1: Videos and their level of description

Fully annotated	**Partly Annotated**	**Not Annotated**
Exam Cheat	Puss in Boots	House of Lies Season 4 Ep 11
London Has Fallen	House of Lies Season 4 Ep 6	House of Lies Season 4 Ep 12
Minions 2015	House of Lies Season 4 Ep 7	Music
Rowan Atkinson Live	House of Lies Season 4 Ep 8	Sherlock Season 2 Ep 3-1
Swimming Pool	House of Lies Season 4 Ep 9	Sherlock Season 2 Ep 3-2
Mr. Bean Back to school	House of Lies Season 4 Ep 10	Suicide Squad
Bratz Desert Jewelz	Mr. Bean Car Park Chaos	
Sherlock Season 2 Ep 1-1	Batman vs. Superman	

Sherlock Season 2 Ep 1-2	On a plane with Mr. Bean	
House of Lies Season 4 Ep 1	Brave	
House of Lies Season 4 Ep 2	Sherlock Season 2 Ep 2-1	
House of Lies Season 4 Ep 3	Sherlock Season 2 Ep 2-2	
House of Lies Season 4 Ep 4		
House of Lies Season 4 Ep 5		
Mr. Bean Getting up late for the dentist		

Table 5-1 above highlights the different videos we used during this research and the extent to which they were annotated with metadata. The MCDMS allows the user to upload a video as long as one or more of the metadata elements given on the GUI are supplied. This was done so that we could be able to evaluate the difference in retrieval between the different levels of video annotations.

5.2 The Results

In this section, we present the results of the testing described in section 5.1. Discussion and interpretation of the results follow in section 5.3.

5.2.1 Smoke test of individual components

- The errors obtained from the smoke tests were mostly issues related to version upgrades and compatibility with the OS running on the PC.

- The interrogation of videos using FFprobe did not always return all the elements expected, e.g. the duration of the video that has just been probed would not be returned and there would be N/A in place of the duration.
- The most common errors when probing were ***No such file or directory*** and ***Argument 'XXX' provided as input filename, but 'YYY' was already specified.***
- The MPEG-7 Library kept returning an error about the lack of Xerces.
- The MPEG-7 returned an error about failure to link to shared libraries.
- Version 2.7.1 of the MPEG-7 Library is based on MPEG-7 2009 DAVP hence it kept returning an error about failing to locate the Profile.
- The GUI managed to connect to the Sedna database effectively.
- The GUI accepted the addition and deletion of metadata elements into its text fields.

5.2.2 Big Bang test of combined components

- FFprobe could not write its results directly into the GUI and that required the user to enter the results from the probe into the GUI test fields manually.
- The FFprobe component did not activate immediately when the user selected describe video on the Video Description Form. It required the video to be uploaded/selected and added by using the link on the Description tab before it activated.
- The Description Module functioned effectively when some of the instructions were carried out manually, e.g. probing the video then adding the metadata elements manually.
- There was the seamless creation of metadata from the description assigned by the user to the video.
- The Description Module carried out its functionality effectively and a success dialog popup appears when the description has been submitted.

- Search for the uploaded videos was slower in terms of the time it took to return results compared to the time it took to describe a video and upload it on the MCDMS.
- Querying the Sedna database seemed to take longer as a connection had to be established.
- Search and retrieval of video content that has been annotated with metadata in a standard MPEG-7 manner proved to be much more efficient than the retrieval of video content without adequate description (annotation)

5.3 Discussion of Results

In this section, the results presented in section 5.2 are discussed in detail.

- The common belief when FFprobe does not return certain metadata elements from its interrogation would be that the video file is corrupted. However, it can always be solved by fully decoding the file using the command: **ffmpeg –i input.filename –f null –**
- Fully decoding the video will return the results required, however, that will require the user to manually probe the video via the terminal using their own command and not using the VideoInterrogation text file to retrieve the technical information about the video being uploaded. It should be noted however that this may take some time.
- The MPEG-7 Library needs to be linked against the libraries libxerces-c and libMp7Jrs. Without these libraries being linked to the MPEG-7 library, the functionality of the MPEG-7 library is disabled as it cannot carry out its serializing (or de-serializing) and validating functions. The library also uses the DAVP which is an upgrade of the Audiovisual Profile (AVP). The DAVP was formulated in 2009 hence the namespace changes and this requires the user to set the namespace from **xmlns:mpeg7="urn:mpeg:mpeg7;schema:2004"** to

xmlns:mpeg7="urn:mpeg:mpeg7;schema:2009"

- The MPEG-7 Library version 2.7.1 that we used for the development of the MCDMS requires Xerces 3.1.1. The application using the MPEG-7 Library must link against the same version of Xerces. The library assumes that the Xerces root directory is on the same level as the MPEG-7 library root directory, hence the error about the lack of Xerces when they are not on the same level. An error about the lack of Xerces can also be returned if the Xerces version used is not version 3.1.1 as stipulated by the library requirements.
- The search and retrieval of video content that had been fully annotated were faster and more efficient in comparison to the search and retrieval of less annotated video content. The efficiency is a result of the finer granularity of search performed. In simpler terms, the more metadata to compare a video against, the smaller the number of videos that will be returned because a finer description is given. An example was the search of the series "House of Lies Season 4 Ep 2" in comparison with "House of Lies Season 4 Ep 11". The search results of episode 11 returned 12 videos, House of Lies Season 4 Ep 1-12 whereas, the search for episode 2 returned just 5 episodes. Episode 11 did not have a lot of metadata annotated to it hence it retrieved all the entries with "House of Lies".
- The GUI did not have any errors returned with regards to invalid entries since it made use of text fields and text-areas only. The GUI allows the user to submit content without describing it so that we are able to compare the difference between searching for fully-annotated, partly-annotated, and non-annotated content altogether. We discovered that the more metadata appended the better the search results. However, in real practice, it is not recommended to allow the system to save a document without annotations. Therefore it is will be wise to later upgrade the MCDMS GUI to avoid such entries.

- Another concern was the issue of tag formatting from the FFprobe results which we later removed using the **-pretty** command and the other command we could have used was the **print_format flat** command.

5.4 Summary

This chapter focused on the functional testing undertaken in this work, the results obtained from the tests, and then went on to give a discussion on the results obtained. The chapter highlighted the smoke tests undertaken and how the Big Bang approach was taken in testing the combined system components in relation to each other. It brought to light how the amount of metadata appended to video content affects the search and retrieval of such-named video. The following chapter gives the conclusions made from the research and highlights the future work.

6. Chapter 6: Conclusions and Future Work

This closing chapter presents a discussion of the findings of the research in parallel with the research objectives presented in the first chapter. It aims to show how the research findings tally with the research questions outlined, stating the extent to which the research statement has been fulfilled by providing the main achievements and conclusions of the research. A comprehensive summary of the important results of the research is offered. Furthermore, the chapter describes areas of possible further exploration for this study. This chapter concludes by giving an overall conclusion of the research.

6.1 Specific Conclusions

This section addresses the various objectives as set out in Chapter 1 section 1.3 and their conclusions. The structure of this chapter is formulated by the different research goals and their findings. Each objective is recalled, answered and discussed accordingly. We aim to conclude on the research that was undertaken and the research goals stated in section 1.3 guide our conclusions.

6.1.1 Addressing the first objective of the research

- To investigate issues related to metadata generation and retrieval.

The research came up with an extensive review of multimedia content annotation tools. Both MPEG-7 compatible and non-compatible systems were reviewed in Chapter 2. An investigation into these various tools on offer brought forth an understanding of the issues that are largely related to metadata annotation and search and the following is an indication of some of the problems:

- The existence of numerous metadata description standards, each with their own specific purpose poses a nuisance when it comes to which standard to select. The standards themselves vary in terms of complexity and compatibility which deters users from using any of the prescribed metadata description standards hence increasing the amount of non-standardized multimedia content being generated by prosumers.

- The semantic gap has proven to be the bane of multimedia content description since machines do not understand the semantics that humans assign to a particular multimedia content' s description. This has made the attempts to fully automate multimedia content description and retrieval near impossible.
- The vocabulary problem
- The encoding specificity problem
- The lack of technical ability of developers and users alike to create tools that can describe multimedia content adequately

6.1.2 Addressing the second objective of the research

- To investigate the important services of IPTV and the innovative services that could be created for IPTV consumption.

  The research came up with an extensive list of IPTV services that have been listed in section 2.1.3. The list of innovative services that can be created is infinite if we consider the fact that the IPTV platform is open and can be accessed by anyone subscribed to IPTV. A list of some of the value-added innovative services is also presented in section 2.1.3 in Chapter 2.

6.1.3 Addressing the third objective of the research

- To develop a MPEG-7 based multimedia content description and retrieval tool.

  An amalgamation of incremental development and re-use oriented development was used to come up with a prototype for the description and retrieval of video content.

The MCDMS that was developed is explained in detail in Chapter 4 of this thesis. The MCDMS managed to achieve the desired functionality of multimedia content (video) description and retrieval as highlighted in Chapter 5.

6.1.4 Addressing the fourth objective of the research

- To conduct functional testing of the tool developed in this research.

The prototype developed in this research underwent functional testing (individual components, integration, and system testing) in an attempt to prove that the MCDMS can describe multimedia content by annotating video content with descriptions that were converted into XML MPEG-7 descriptions and stored in an XML database. The tests undertaken have been highlighted in section 5.1 of Chapter 5.

6.2 Recommendations and Future Work

As can be said of almost every research that has been carried out, progress will always lead to more questions. We would want to create a testbed by modifying the UCT IMS testbed in relation to the Open IMS stack and testing our tools' compatibility with the IPTV platform. It would also be beneficial to check if other standards could produce better results or if they could be easier to use in comparison to the MPEG-7 standard.

6.3 Overall Conclusions

Search and retrieval of multimedia content has long been a problem and continues to be one. The availability of numerous metadata description standards that can be used for the creation of metadata used to describe the multimedia content does not aid the novice users of such technologies much. There is a need for the users to amass expertise for them to be able to use the existing standards. This research sought to create a tool that implements one such metadata description standard, namely, the Multimedia Content Description Interface (MPEG-7) standard. The MCDMS prototype created in this research proved that the implementation of the MPEG-7 standard by a tool which is easier to use for end-users helps in eradicating the metadata problem described in Chapter 2 of this thesis. The creation of metadata becomes a simpler feat to take on and thus producing detailed descriptions of multimedia content that is uploaded onto open platforms, particularly the IPTV platform which was the focus of this research. It has been proven that the more description a video has the better it is to search and retrieve the video from the video database. Therefore, it goes to prove that the implementation of such a tool as the MCDMS for any of the metadata description standards could be a solution to the metadata problem regardless of what type of media is being described. However, it was noted that the complexity of metadata description standards is not the only problem that heightens the metadata problem. The semantic gap was noted as being one of the huge contributors to the metadata problem as it affects the creation of metadata itself. The research produced a number of issues that affect metadata creation and we hope the issues highlighted can help prosumers of multimedia content with the description of multimedia content when they create content or when they search for the multimedia content they require.

References

Adam, N., Yesha, Y., Awerbuch, B., Bennet, K., Blaustein, B., Brodsky, A., Chen, R., Dogramaci, O., Grossman, B., Holowczak, R., Johnson, J., Kalpakis, K., McCollum, C., Neches, A.L., Neches, B., Rosenthal, A., Slonim, J., Wactlar, H., Wolfson, O. and Yesha, Y., 1996. Strategic Directions in Electronic Commerce and Digital Libraries : Towards a Digital Agora. 28(4), pp.818–835.

ATIS IPTV Research Group, 2005. *ATIS IPTV Exploratory Group Report and Recommendation to the TOPS Council*.

Avison, D.E. and Fitzgerald, G., 2006. *Information Systems Development: Methodologies, Techniques and Tools*. 4th ed. McGraw-Hill.

Bargeron, D., Gupta, A., Grudin, J. and Sanocki, E., 1999. Annotations for streaming video on the Web: System design and usage studies. *Computer Networks*, 31(11), pp.1139–1153.

Bastan, M., Cam, H., Gudukbay, U. and Ulusoy, O., 2010. Bilvideo-7: An MPEG-7-Compatible Video Indexing and Retrieval System. *Multimedia, IEEE*, [online] pp.62–73. Available at: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5692184>.

Bastan, M., 2010. *BilVideo-7 : Video Parsing , Indexing and Retrieval*. Bilkent University.

Bathaee, H., 2014. *Next Generation Network Concepts*. [online] Available at: <http://www.slideshare.net/HavarBathaee/ngn-nextgeneration20network> [Accessed 20 Feb. 2017].

Beal, V., 2015. *IP Multimedia Subsystem*. [online] Available at: <http://www.webopedia.com/TERM/I/IP_Multimedia_Subsystem.html> [Accessed 23 Nov. 2015].

Centers for Medicare & Medicaid Services, 2008. *Selecting a development approach*. *Centers for Medicare & Medicaid Services*, Available at: <http://www.cms.gov/Research-Statistics-Data-and-Systems/CMS-Information-Technology/XLC/Downloads/SelectingDevelopmentApproach.pdf>.

Chang, S., Sikora, T. and Puri, A., 2001. Overview of the MPEG-7 Standard. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on MPEG-7*, (June).

Döller, M. and Lefin, N., 2007. Evaluation of available MPEG-7 Annotation Tools.

Dublin Core Initiative, 2013. *DublinCore Metadata Initiative*. [online] Available at: <http://dublincore.org/> [Accessed 7 Aug. 2015].

ETSI Technical Committee TISPAN, 2007. *ETSI TS 102 822-2: 'Broadcast and On-line Services: Search, select and rightful use of content on personal storage systems*

('TV-Anytime'); Part 2: System description'.

ETSI Technical Committee TISPAN, 2009. *ES 282 001 - V3.4.1 - Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Functional Architecture*.

Friedland, G. and Jain, R., 2014. *Multimedia Computing*. Cambridge University Press.

Google, 2017. *Search by Image – Inside Search – Google*. [online] Available at: <https://www.google.com/intl/es419/insidesearch/features/images/searchbyimage.html> [Accessed 11 Jan. 2017].

IEEE-SA Standards Board, 2000. IEEE Recommended Practice for Architectural Description of Software-Intensive Systems. *IEEE Std*, [online] 1471-2000, pp.1–23. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.102.9904>.

Institute for System Programming RAS, 2012. *Sedna XML Database*. [online] Available at: <http://sednadb.com/> [Accessed 14 Feb. 2017].

ITU-T Newslog, 2006. *ITU-T Newslog - IPTV Standardization on Track Say Industry Experts*. [online] Available at: <http://www.itu.int/ITU-T/newslog/IPTV+Standardization+On+Track+Say+Industry+Experts.aspx> [Accessed 11 Jan. 2017].

ITU-T, 2015. *ITU-T's Definition of NGN*. [online] Available at: <http://www.itu.int/en/ITU-T/gsi/ngn/Pages/definition.aspx> [Accessed 23 Nov. 2015].

Jambor, T., 2012. Intelligent media indexing and television recommender systems. In: *The 3rd BCS IRSG Symposium on Future Directions in Information Access*. pp.50–55.

Janevski, T., 2014. *NGN Architectures, Protocols and Services - Toni Janevski - Google Books*. [online] Wiley Publishers. Available at: <https://books.google.co.za/books?id=ZjEKAwAAQBAJ&printsec=frontcover&dq=inauthor:%22Toni+Janevski%22&hl=en&sa=X&ved=0ahUKEwjt2N3pj7rRAhVCyyYKHXnVAlAQ6AEIGjAA#v=onepage&q&f=false> [Accessed 11 Jan. 2017].

Joanneum Research Institute of Information Systems and Information Management, 2009. *MPEG-7 Detailed Audiovisual Profile (DAVP).*

Joanneum Research Institute of Information Systems and Information Management, 2013. *MPEG-7 C++ API Implementation*.

José M. Martínez (UAM-EPS-GTI, E., 2004. *MPEG-7 Overview*. [online] Palma de Mallorca. Available at: <http://mpeg.chiariglione.org/sites/default/files/files/standards/docs/w6828_mp7_Overview_v10.zip>.

JSB Market Research, 2014. *Emerging Markets Lead IPTV Subscription Growth Through 2018. Research in Focus: IPTV*. [online] Available at: <http://www.slideshare.net/JSBMarketResearch/emerging-markets-lead-iptv> [Accessed 14 Feb. 2017].

Kaplan, J., 2016. *Trump Archive*. [online] Available at: <http://archive.org/details/trumparchive?sort=titleSorter> [Accessed 17 Jan. 2017].

Kothari, C.R., 2004. *Research Methodology : Methods & Techniques*. New Age International (P) Ltd.

Library of Congress, 2016. *Metadata Encoding and Transmission Standard (METS) Official Web Site | Library of Congress*. [online] Available at: <http://www.loc.gov/standards/mets/> [Accessed 11 Jan. 2017].

Lindgren, P., 2010. *Diameter service creation investigation and HSS evaluation Diameter service creation investigation and HSS evaluation*. KTH Royal Institute of Technology.

Magalhaes, J.M. da C., 2002. *Universal Access to Multimedia Content Based on the MPEG-7 Standard*. [online] *Access*. Universidade Tecnica De Lisboa. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.6810&rep=rep1&type=pdf>.

Marcos, G., Illarramendi, A., Olaizola, I.G. and Flórez, J., 2011. *A Semantic Middleware to enhance current Multimedia Retrieval Systems with Content-based functionalities*. *Multimedia Systems*. Universidad del Pais Vasco/ Euskal Herriko Uribertsiatea.

Mikoczy, E., Sivchenko, D., Xu, B. and Rakocevic, V., 2007. IMS based IPTV services: architecture and implementation. *Proceedings of the 3rd International Conference on Mobile Multimedia Communications*, [online] (November), p.15. Available at: <http://dl.acm.org/citation.cfm?id=1385289.1385307>.

National Information Standards Organization, 2004. *Understanding Metadata*. [online] *National Information Standards*, Available at: <http://www.mendeley.com/catalog/understanding-metadata-3/\nhttp://www.niso.org/publications/press/UnderstandingMetadata.pdf>.

Ncube, P.D.N. and Shibeshi, Z.S., 2015. Design of an MPEG-7 based Multimedia Content Description and Management System (MCDMS) for Internet Protocol Television (IPTV). In: *Southern Africa Telecommunications Networks and Applications Conference (SATNAC), in press*.

NISO Framework Working Group, 2007. *A Framework of Guidance for Building Good Digital Collections*. 3rd Editio ed. [online] *2009-06-28]*. *Http://Www. Niso. Org/Publications/*, Available at: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Framework+of+Guidance+for+Building+Good+Digital+Collections#2>.

Pereira, F., 2000. *‘MPEG-7 Requirements’, ISO/IEC JTC1/SC29/WG11/N3933*. [online] Available at: <https://www.mendeley.com/library/#> [Accessed 11 Jan. 2017].

Perperis, T., Tsekeridou, S. and Theodoridis, S., 2007. An Ontological Approach to Semantic Video Analysis for Violence Identification. In: *Proceedings of I-MEDIA ’07 and I-SEMANTICS '07*. Graz, Austria, pp.139–146.

Project Isizwe, 2015. *Home Page - Project Isizwe*. [online] Available at: <http://www.projectisizwe.org/> [Accessed 20 Feb. 2017].

RDF Working Group, 2014. *RDF - Semantic Web Standards*. [online] Available at: <https://www.w3.org/RDF/> [Accessed 21 Feb. 2017].

Singh, I., 2015. *Fixed Mobile Convergence*. [online] Available at: <http://www.slideshare.net/InderBarara1/fixed-mobile-convergence-47620069> [Accessed 20 Feb. 2017].

Snoek, C.G.M. and Worring, M., 2007. Concept-Based Video Retrieval. *Foundations and Trends® in Information Retrieval*, [online] 2(4), pp.215–322. Available at: <http://www.nowpublishers.com/article/Details/INR-014> [Accessed 11 Jan. 2017].

Sommerville, I., 2000. *Software Engineering Sixth Edition*. 6th ed. [online] Addison Wesley Publishing Company. Available at: <http://www.powells.com/book/software-engineering-6th-edition-9780201398151> [Accessed 13 Jan. 2017].

Sommerville, I., 2011. *Software Engineering*. 9th ed. Pearson.

Sommerville, I., 2015. *Software Engineering (10th Edition)*. 10th ed. [online] Pearson. Available at: <https://www.amazon.com/dp/0133943038/ref=rdr_ext_tmb> [Accessed 13 Jan. 2017].

Spaniol, M., Klamma, R. and Lux, M., 2008. Content Based Retrieval & Beyond. 14(10), pp.1792–1807.

Steves, M.P., Ranganathan, M. and Morse, E., 2001. SMAT: Synchronous Multimedia and Annotation Tool. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, 00(c), pp.1–10.

Takeda, H., Veerkamp, P., Tomiyama, T. and Yoshikawa, H., 1990. Modeling Design Processes. *AI Magazine*, 11(4), pp.37–48.

Techopedia, 2017. *What is Functional Testing? - Definition from Techopedia*. [online] Available at: <https://www.techopedia.com/definition/19509/functional-testing> [Accessed 15 Feb. 2017].

Text Encoding Initiative, 2016. *TEI: Text Encoding Initiative*. [online] Available at: <http://www.tei-c.org/index.xml> [Accessed 11 Jan. 2017].

The Apache Software Foundation, 2016. *The Apache Xerces^{TM} Project - xerces.apache.org*. [online] Available at: <http://xerces.apache.org/> [Accessed 14 Feb. 2017].

The Moving Pictures Experts Group, 2016. *The Moving Pictures Experts Group: MPEG-7*. [online] Available at: <http://mpeg.chiariglione.org/standards/mpeg-7> [Accessed 7 Aug. 2015].

Tochtermann, K., Haas, W., Kappe, F., Scharl, A., Pellegrini, T. and Schaffert, S., 2007. *I-MEDIA '07, I-SEMANTICS '07: International Conferences on New Media Technology and Semantic Systems*.

Tsai, J.-W., Huang, C.-C., Chang, C.-S., Tang, C.-S. and Twu, C.-Y., 2010. The technique of IMS Sh interface and its NGN LBS application. In: *International Conference on Communication Technology Proceedings, ICCT*. pp.599–602.

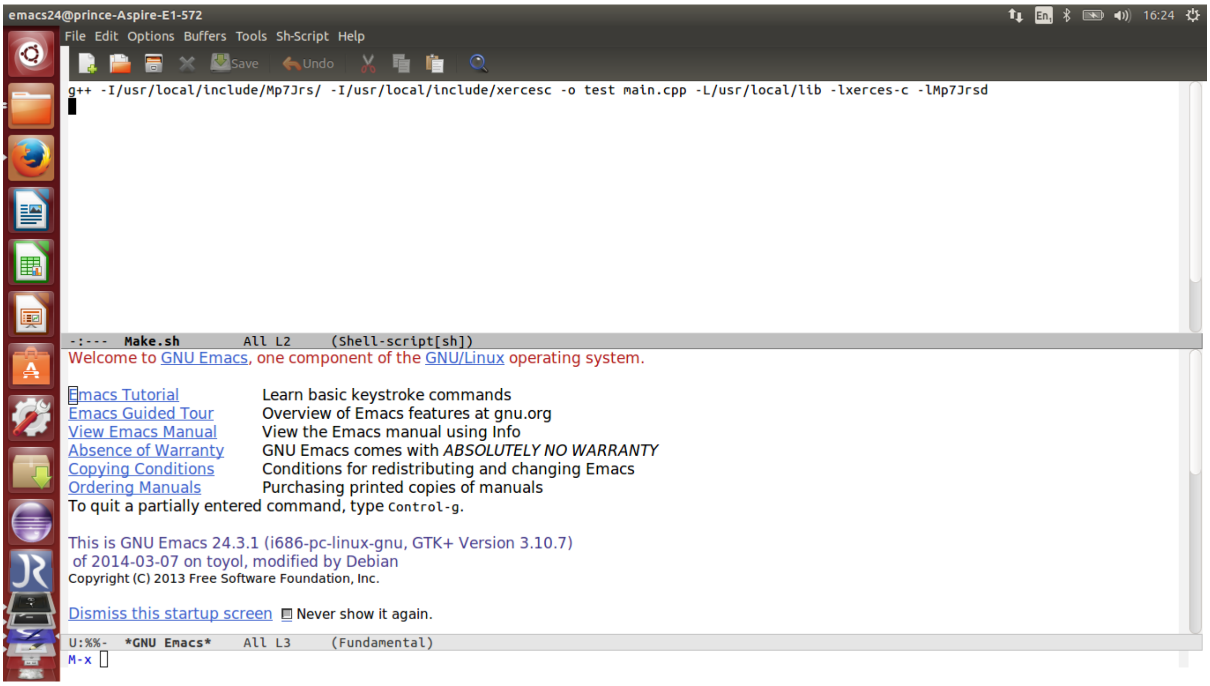
Versatile Delivery Systems Ltd, 2006. *Frameline 47 uses MPEG-7 for Descriptive Data for Videos*. [online] Available at: <http://www.geniusdv.com/weblog/archives/defunct/frameline_47_uses_mpeg7_for_descriptive_data_for_videos.php>.

Vries, A.P. De, 1999. *Content and Multimedia Database Management Systems*. *Technology*. University of Twente.

YouTube, 2015. *YouTube Statistics*. [online] Available at: <https://www.youtube.com/yt/press/statistics.html> [Accessed 31 Aug. 2015].

Zhao, R. and Grosky, W.I., 2002. Narrowing the semantic gap - improved text-based web document retrieval using visual features. *IEEE Transactions on Multimedia*, 4(2), pp.189–200.

Appendix A. Linking to the libraries libxerces-c and lMp7Jrsd



Appendix B. The Simple Program for creating a description

```
#include <stdio.h>
#include <xercesc/util/PlatformUtils.hpp>

#include "Mp7JrsEverything.h"
#include "Mp7JrsArchive.h"
#include "Mp7JrsException.h"

int main(int argc, char** argv)
{
	// Initialise Xerces
	XMLPlatformUtils::Initialize();
	// Initialise MPEG-7 JRS DAVP
	Mp7JrsFactory::Initialize();

	Mp7JrsArchive archive;

	archive.SetEncoding("UTF-8");
	archive.SetDoValidation(true);
	archive.SetSchemaLocation("urn:mpeg:mpeg7:schema:2004", "mpeg7",
"davp-2009.xsd");

	try
	{
```

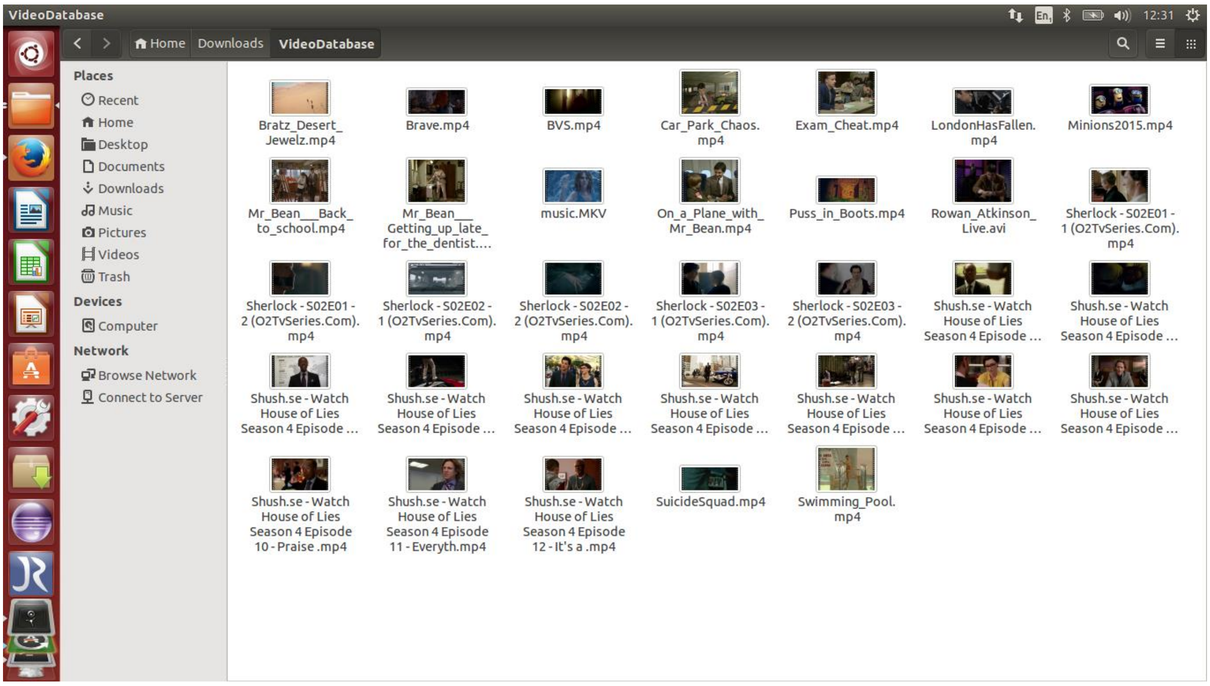
```
        Mp7JrsNodePtr root = archive.FromFile("test.xml");
    }
    catch(Mp7JrsException & e)
    {
        const char * msg = e.GetMessageText();
        printf("ERROR: %s\n", msg);
    }
    catch(...)
    {
        printf("ERROR: %s\n", "Unknown error");
    }
    // Cleanup MPEG-7 JRS DAVP
    Mp7JrsFactory::Terminate();
    // Cleanup Xerces
    XMLPlatformUtils::Terminate();

    return 0;
}
```

Appendix C. The Validate Program

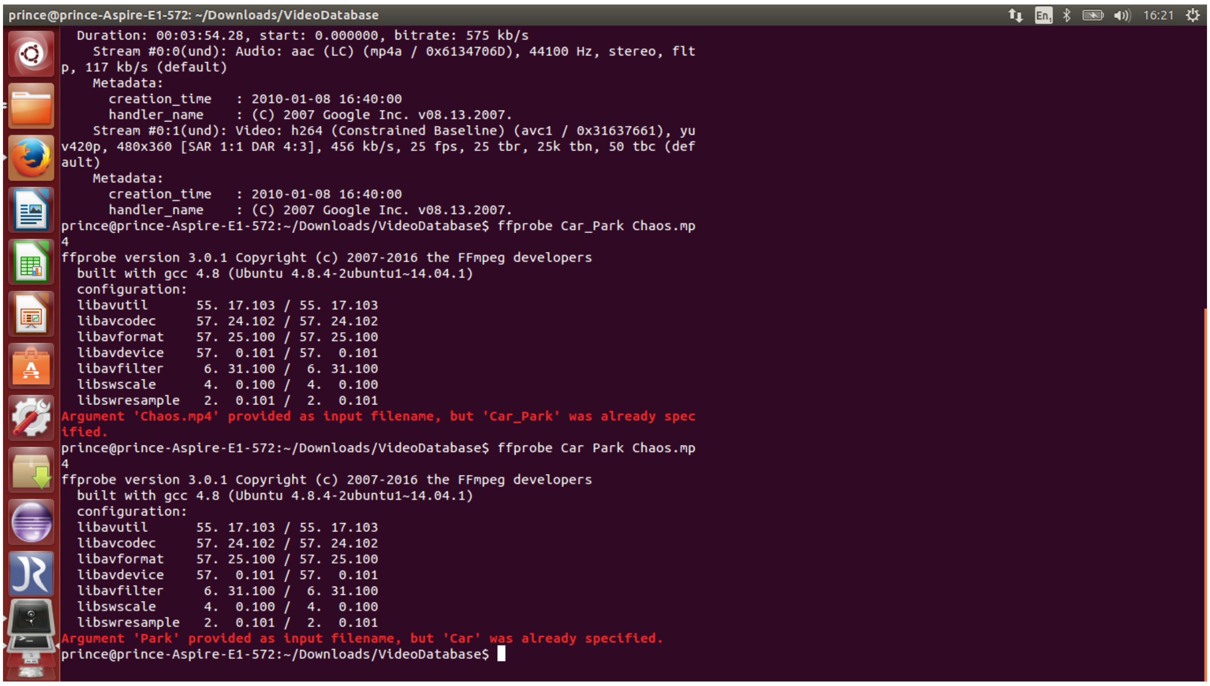
```
#include <stdio.h>
#include <xercesc/util/PlatformUtils.hpp>

#include "Mp7JrsEverything.h"
#include "Mp7JrsArchive.h"
#include "Mp7JrsException.h"

int main(int argc, char** argv)
{
	// Initialise Xerces
	XMLPlatformUtils::Initialize();
	// Initialise MPEG-7 JRS DAVP
	Mp7JrsFactory::Initialize();

	Mp7JrsArchive archive;

	archive.SetEncoding("UTF-8");
	archive.SetDoValidation(true);
	archive.SetSchemaLocation("urn:mpeg:mpeg7:schema:2004", "mpeg7",
"davp-2009.xsd");

	try
	{
```

```
        Mp7JrsNodePtr root = archive.FromFile("test.xml");
    }
    catch(Mp7JrsException & e)
    {
        const char * msg = e.GetMessageText();
        printf("ERROR: %s\n", msg);
    }
    catch(...)
    {
        printf("ERROR: %s\n", "Unknown error");
    }
    // Cleanup MPEG-7 JRS DAVP
    Mp7JrsFactory::Terminate();
    // Cleanup Xerces
    XMLPlatformUtils::Terminate();

    return 0;
}
```

Appendix D. Videos Used



Appendix E. Common Errors



```
prince@prince-Aspire-E1-572: ~/Downloads/VideoDatabase
  Duration: 00:03:54.28, start: 0.000000, bitrate: 575 kb/s
    Stream #0:0(und): Audio: aac (LC) (mp4a / 0x6134706D), 44100 Hz, stereo, flt
p, 117 kb/s (default)
    Metadata:
      creation_time   : 2010-01-08 16:40:00
      handler_name    : (C) 2007 Google Inc. v08.13.2007.
    Stream #0:1(und): Video: h264 (Constrained Baseline) (avc1 / 0x31637661), yu
v420p, 480x360 [SAR 1:1 DAR 4:3], 456 kb/s, 25 fps, 25 tbr, 25k tbn, 50 tbc (def
ault)
    Metadata:
      creation_time   : 2010-01-08 16:40:00
      handler_name    : (C) 2007 Google Inc. v08.13.2007.
prince@prince-Aspire-E1-572:~/Downloads/VideoDatabase$ ffprobe Car_Park Chaos.mp
4
ffprobe version 3.0.1 Copyright (c) 2007-2016 the FFmpeg developers
  built with gcc 4.8 (Ubuntu 4.8.4-2ubuntu1~14.04.1)
  configuration:
  libavutil      55. 17.103 / 55. 17.103
  libavcodec     57. 24.102 / 57. 24.102
  libavformat    57. 25.100 / 57. 25.100
  libavdevice    57.  0.101 / 57.  0.101
  libavfilter     6. 31.100 /  6. 31.100
  libswscale      4.  0.100 /  4.  0.100
  libswresample   2.  0.101 /  2.  0.101
Argument 'Chaos.mp4' provided as input filename, but 'Car_Park' was already spec
ified.
prince@prince-Aspire-E1-572:~/Downloads/VideoDatabase$ ffprobe Car Park Chaos.mp
4
ffprobe version 3.0.1 Copyright (c) 2007-2016 the FFmpeg developers
  built with gcc 4.8 (Ubuntu 4.8.4-2ubuntu1~14.04.1)
  configuration:
  libavutil      55. 17.103 / 55. 17.103
  libavcodec     57. 24.102 / 57. 24.102
  libavformat    57. 25.100 / 57. 25.100
  libavdevice    57.  0.101 / 57.  0.101
  libavfilter     6. 31.100 /  6. 31.100
  libswscale      4.  0.100 /  4.  0.100
  libswresample   2.  0.101 /  2.  0.101
Argument 'Park' provided as input filename, but 'Car' was already specified.
prince@prince-Aspire-E1-572:~/Downloads/VideoDatabase$
```

Appendix F. FFprobe output on the terminal