FULL-TEXT SEARCH OPTIMIZATION IN DIGITAL LIBRARY ANDROID APPLICATION

By
Arayan, Andrew Rey T.
Elnas, Jeric B.

ATENEO DE DAVAO UNIVERSITY
SCHOOL OF ARTS AND SCIENCES
DAVAO CITY

OCTOBER 2013

FULL-TEXT SEARCH OPTIMIZATION IN DIGITAL LIBRARY ANDROID APPLICATION

A Thesis / Research Project

Presented to

The Faculty of the Computer Studies Division

Ateneo de Davao University

In Partial Fulfillment

Of the Requirements for the Degree

Bachelor of Science in Information Technology

By

Arayan, Andrew Rey T.

Elnas, Jeric B.

SCHOOL OF ARTS AND SCIENCES

ATENEO DE DAVAO UNIVERSITY

OCTOBER 2013

Table of Content

Full-text Search Optimization in Digital Library Android Application

ANDREW REY ARAYAN, Ateneo de Davao University
JERIC ELNAS, Ateneo de Davao University

Database queries are typically specified declaratively, and the database system has to choose an appropriate execution plan for the query. A query optimization in a database system is responsible for transforming a SQL query into an execution plan. It is also responsible for taking a user query and searching through the entire space of equivalent execution plans for a given user query and returning the execution plan with the lowest cost. There are problems that the user encounters when searching in a digital library. The inaccurate results that is irrelevant to the user. Traditional queries swallows large amount of system resources in order to execute a query and even slower results. By integrating Full text search optimization in an information retrieval with the use of full text search optimization, the researchers will be able to test the usability of query optimization in a digital library. This study will also help identify different strengths and weaknesses of query optimization by creating digitized library system. This study discusses the different methodology for the development of digital library android application using HTML5 using JQuery and integrating full text search optimization in information retrieval. The researchers will also seek to create a digitized library system.

General Terms: Query Optimization

Additional Key Words and Phrases: HTML5, Android, Android Application, Digital Library, Query Optimization, Information Retrieval

ACM Reference Format:

Andrew Rey Arayan and Jeric Elnas. 2013. Full-text Search Optimization in Digital Library Android Application

1. INTRODUCTION

1.1 Background of the Study

Smartphones offer a lot of possibilities doing the things that previously required a computer or other technical devices. If your device does not have the application you need, just download it through your wireless internet and start using it instantly. The use of smartphones is totally increasing because of the usability and capability of the technological device. It is a mobile operating system that runs on different platform. Smartphones has the capability of a desktop computer from gaming, video streaming and others. Children, teens and even adults use smartphones on their daily activities.

One of the current technologies that run in an android platform is HTML5. HTML5 is the fifth major revision of core language of the internet which is HTML. The newly introduced HTML5 provides a lot of new possibilities that will help the users in mobile application development by proving many new API's along with other useful features.HTML5, together with CSS3 and JavaScript, could create simple to complex applications. Though HTML5 functionality is being limited to new browsers, it perfectly works on smartphones such as IOS, Blackberry and Android.

Full text search optimization refers to techniques in which it searched in a computer stored document in a database. It examines every word in the stored document as it tries to compare the word that is searched from the database. The full-text application will be down for a considerably shorter time as compared to the full population method. Even with very large full-text catalogs, this method should take no more than a few hours for most large deployments. By applying this technique to search queries, it gives better results and lesser retrieval time.

The goal of information retrieval is facilitating a user's access to information that is relevant to his information needs. Information Retrieval systems provide users easy access to the information where the person is interested. Through the use of query optimization, it reduces the system resources required to fulfill a query, and it provides the user with the correct result set faster. There are several reasons why query optimization is important. First, it provides the user with faster results, which makes the application seem faster to the user. Second, it allows the system to execute more queries in the same amount of time, because every request takes lesser time than queries that are not optimized. And last, query optimization reduces the amount of wear on the hardware, and allows the server to run more efficiently.

Because of the growing demand of a digital library, software developers are now trying to develop a better application for digital libraries. One way to have a good application on digital library is by applying query optimization. By applying query optimization in information retrieval by developing a digital library, this will help application to have faster result rather than using the traditional query techniques. And last, applying query optimization reduces the amount of the system resources that are consumed by the query.

1.2 Technology Application Context

The present study seeks to the development of digital library in an android application using HTML5. In developing digital library in an android platform, integrating it with Full text search optimization will greatly help to have faster and even accurate results. By Full text search optimization in digital library, system resources can be minimized and will perform even better in contrast with the traditional query techniques. Through this, the researchers will be able to identify different strengths and weaknesses Full text search optimization when applied to digital library android application. To identify the usability of query optimization in terms of its performance and how much system resources does it take to perform a query.

After studying different query optimization techniques and its problems, the researchers found out that there are different problems in query optimizations. One the problem was on how to integrate information retrieval to query optimization aside from using Boolean search techniques. Another problem was that the researchers found out were the imprecise behavior when integrating information retrieval to DBMS after applying query optimizations. Traditional queries swallows larges amount of system resources when executing a query which leads to lag or slower searching results. And last problem is the inaccurate result which is irrelevant to the users. Despite this problems, there are query optimization algorithms that will help solve this problems and still applying query optimization greatly helps the information retrieval and rather than using the traditional query techniques.

1.3 Objectives of the Study

The general objective of this research is to develop an android application using HTML5 and applying Full text search query optimization in digital library. The specific objectives are:

- To identify the strength and weaknesses of Full text search optimization when applied to a digital library.
- To identify the usability of Full text search optimization when applied in a digital library
- To identify the accuracy and speed of Full text search optimization when searching in a digital library.
- To developed an optimized digital database library system

1.4 Significance of the Study

This study is significant for all smartphone users using the android platform connected to Wi-F or mobile data. Through a digital library that will search and borrow for books, Android device users

will be able to access and view more books without getting into the hassle of going to university library for reservation and borrowing of books.

This thesis gives deeper understanding of HTML5 deployment scenario in android platform, browser deficiency or dependency and users experience while accessing the application through the use of smartphones.

This study also contributes to information retrieval through the use of Full text search optimization. Full text search optimization is one of the most efficient way in an information retrieval. Applying Full text search optimization in an information retrieval by developing a digital library in an android platform reduces the system resources required to fulfill a query. This provides the users with accurate and faster result.

1.5 Scope and Limitations of the Study

The study will focus on creating digital library in an android Platform. Android is chosen to use in the study because statistics shows that it is one of the most used platform in smartphones. The study will also on full text query optimization that will use in the search query. The researchers will test the effectiveness, capability and usability of query optimization, full text search optimization, when applied in a digital library android platform.

The researchers will use other optimization techniques that will be used to compare to full text search optimization by using different existing benchmarking tools. One technique that will be used to would be the traditional type of query optimization. It is the most basic type of technique that is applied in search queries. And the other technique is the like optimization technique because it has similar characteristics to full text search optimization.

2. REVIEW OF RELATED LITERATURE AND TECHNOLOGY APPLICATIONS

Prior to the Development of Digital Libraries in an android application there are some research conducted which uses same tools and related techniques that are essential on developing the digital libraries.

A study conducted by Sirisha Jonnalagadda (2012) in developing digital library in San Diego State University. In his study, he created a digital library that has all the features that the user would need when he/she visits the library website which includes a one-stop solution for retrieving all the library related information and others. There are two primary features of his digital library. First, the application allows user to search in library's catalog based on the entered keywords. And the other is the bibliographic information of a particular book along with its status which includes its availability, location and call number. However, there is a problem to this study. Sirisha Jonnalagadda pointed out that since the application fetches information from the WorldCat Search API and not from the local servers directly, there are instances that the two databases would not sync.

Another study was conducted by Van Vught (2002) regarding on database optimization aspects for information retrieval. In his study, he cited the different problems on information retrieval when combining it t query optimization. On the problem was the imprecise or inexact query processing. When integrating DBMS with IR, the DBMS needs to be told on how to deal with this imprecise behavior. There are many proposed solution for this problem, but, they will be using the traditional approach or using Boolean search which nowadays now often consider it as too crude on capturing its impreciseness. Another problem that was pointed out was that IR algorithm was designed to operate on the data stored in a specific way. Thus, this algorithm conflicts with the data independence property of a typical DBMS approach. So this algorithm therefore is clearly not beneficial to a better performance. And the last problem that was cited was IR query processing actually is an alternating sequence of queries and feedback, in contrast with database query processing which typically assumes ad hoc queries. So in this case, query optimizer should really be aware this relationships between queries in time.

A study made by Raf Walraven (2011) in his thesis, "Realizing Efficient Query Distribution in the Mobile Semantic Web". This study was executed in an android digital library android application, the Jena library. He proposed two optimization techniques for the distribution plan. First, outsourcing joins to external query endpoints in order to increase performance. And second, by maintaining an index, to rule out parts of the query distribution plan that will not yield results. So they index the graph patterns occurring in the data sources.

The result in his first proposed techniques, the process was already slightly optimized, by using a caching mechanism for the results of a subquery, to reduce the amount of requests. In his second proposed optimization technique, the index optimization was said to be the best optimization among the two proposed techniques. Raf Walravens stated that, "The graph patterns in an index, can be matched to the subqueries via graph comparison, since the WHERE-clauses of SPARQL (sub) queries also have an inherent graph structure. Comparison is currently based on predicates' labels and their adjacent predicates."

Sang Ho Lee in his study, "A query Transformer in a Digital Library Environment", discusses about search engines and its different problems that occurred like its irrelevance when searching. In his study, he proposes to use the concept of query transformer in which it refines the user query so that what the user searches in closely relevant to the user's intention, by using semantic query optimization techniques in transforming the query.

Sang Ho Lee discussed on the results of his study on query transformer. He pointed out that using query transformer is easily applicable to a wide range of environments which includes the Web, digital libraries and Usenet. He also discussed that query transformer should be updated over its knowledge base accordingly due to user's search characteristics which are subjected and may change over time.

3. METHODOLOGY

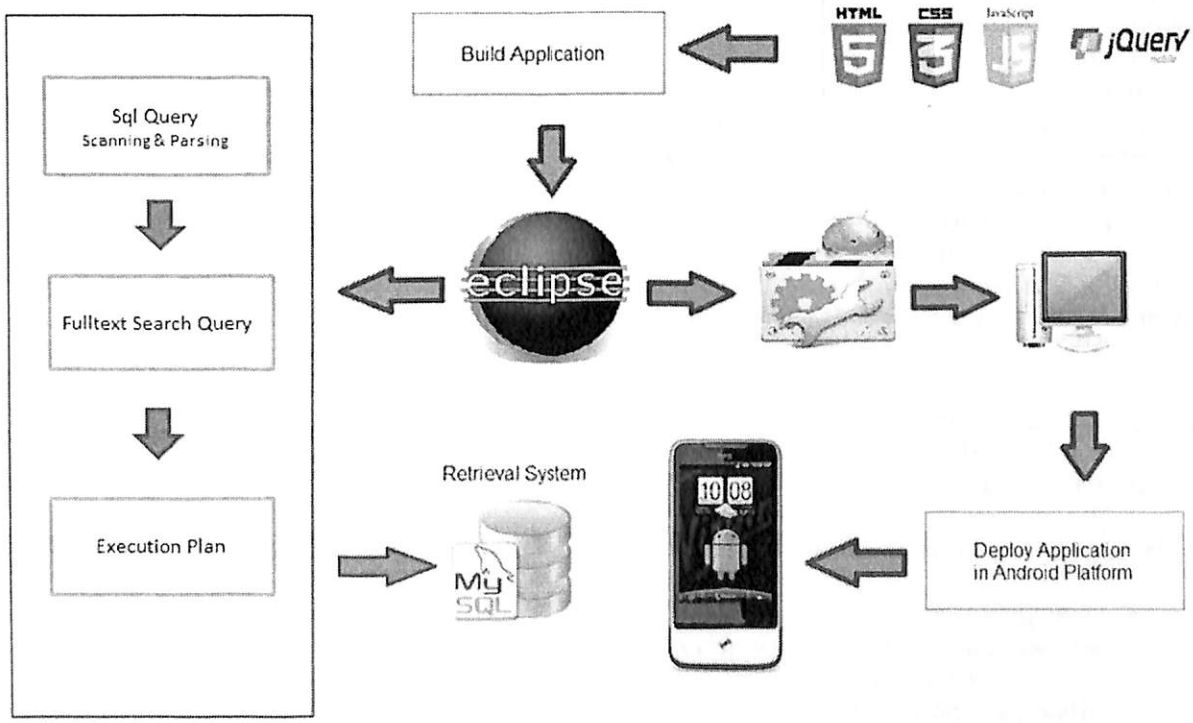


Fig 1 Research/Project design

This study was conducted to provide a Digital Library using the HTML5 in an android platform. The objective of the research is to evaluate the usability and productivity of the query optimization used in retrieving data and to come up with the best execution plan. In order to achieve the goal of this study, researchers choose some published articles and literatures as primary resources. Developing the project, consider different areas. First, to create a simple application that user can understand. The simplicity of the application is important to gain the attention of the user. HTML5 together with the JavaScript and CSS will be used for designing and building a user friendly interface. The researchers will also an open source framework that is compatible with android which is the Jquery mobile. Jquery mobile is a framework that. Since HTML5 is emerging in creating and deploying a smartphone mobile application. Therefore, used for developing the frontend of the project. The recommended environment for developing Android application is Eclipse with the Android Development Toolkit (ADT) plugin installed and with the help of Android SDK which Google provides for free in developing an Android-based smart device application. Next, to gather information's from the university library to create a reliable digital library. Collecting Information of the book in the library and save in the database that will be used for information retrieval. And lastly, to choose a best query plan to create the best execution that can retrieve information with the lowest cost. Researchers test some researched query optimization for getting the pacing execution that will be implemented to the project.
Benchmarking tools will be used to test the search queries. One of the benchmarking tools that will use is dbForge. dbForge is a benchmaking tool that optimized the queries and at the same time it test if the queries are optimized. The time that it takes in retrieving the queries is the bases that the query is an optimized one.

4. TECHNOLOGY BACKGROUND

4.1 Query Optimization

Query optimization born since before 1970s. those times were called the dark ages because query optimizations are done manually. Also, the relational database is not yet known during those times so it requires a person that is really an expert in the database to perform query optimization.
As the time goes by, so with the technology that becomes advanced. In the middle of 1970s and 1980s, query optimization is no longer done manually, it is done by a system known as R systems that later evolved into optimization command JOIN which sets the stage for the next query optimization. Until now, query optimization continues to grow together with the emergence of new techniques used in the query optimization process. It is said that query optimization is the backbone of the system because the effectiveness and efficiency of the system depends on how it was optimize.
Almost all database management systems using a query optimization system to decrease the query execution time is optimized so that the system works and use of the resources can be minimized.

4.2 Types query optimization algorithms

a. Semijoin Based Algorithm - The semijoin acts as a size reducer (similar as to a selection) such that smaller relations need to be transferred. The semijoin is beneficial if the cost to produce and send it to the other site is less than the cost of sending the whole operand relation and of doing the actual join.
b. INGRES Algorithm - uses a dynamic query optimization algorithm that recursively breaks a query into smaller pieces.
c. Distributed INGRES Algorithm - the distributed one should break up each query qi into subqueries that operate on fragments; only horizontal fragmentation is handled. Optimization with respect to a combination of communication cost and response time.
d. System R Algorithm - Performs static query optimization based on "exhaustive search" of the solution space and a cost function.

e. Hill-Climbing Algorithm - Refinements of an initial feasible solution are recursively computed until no more cost improvements can be made.

There are six different techniques or methods in using query optimization. First, use index. Using an index will help to speed up a query. Second, use aggregate table. Pre-populating tables at higher levels so fewer amounts of data need to be parsed. Third, use vertical partitioning so that it will decrease the amount of data a query needs to process. Fourth, using horizontal partitioning will also help decrease the amount of data needs to process. Fifth, denormalization which combines multiple tales into a single table that helps to speed up a query performance because fewer table joins are needed. And last, server tuning. Each tuning has its own parameters, and often tuning server parameters so that it can fully take advantage of the hardware resources that can significantly boost up the query performance.

5. RESULTS AND ANALYSIS

Before developing the application, the proponents went to Ateneo de Davao's college library to have a soft copy of library's list of books. The proponents provided a signed letter of IT chairperson, the proponent's adviser, and the proponents themselves; the proponents were able to obtain a soft copy of list of books available. The list of books will be used in the digital library android application.

Wamp server is used in building up the database which the list of books provided by Ateneo de Davao's college library is stored. The application must be connected to the internet via mobile data connection or Wi-Fi to connect to the server and use the database.

5.1 DESIGN

5.1.1 Building up Digital Library

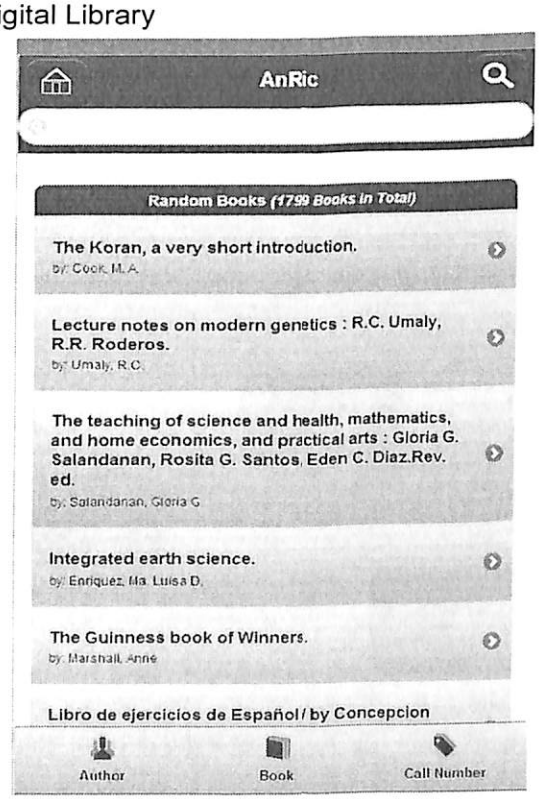


Fig 3 jQuery Mobile platform

In building up the digital library, the proponents used JQuery mobile as a platform because it is a HTML5-based user interface that can be applied on all popular mobile platforms. Using Eclipse, an ADT (android development tool), to open the header file in JQuery and used it to design the user interface. The header sets all the overall design of the application specifically the themes, colors and other important things that are used in designing the application.

```
<head>
<style>
lable { text-align:left; width:120px; display:block; color:Black; float:left; clear:left; }
</style>
  <meta name="viewport" content="user-scalable=no,width=device-width" />
  <meta http-equiv="Content-Type" content="text/html; charset=utf-8">
  <meta name="viewport" content="width=device-width,initial-scale=1,maximum-scale=1,user-scalable=no"/>

  <link rel="stylesheet" href="jquery.mobile/theme.min.css" />
  <link rel="stylesheet" href="jquery.mobile/jquery.mobile.structure.css" />
  <link rel="stylesheet" href="jquery.mobile/jquery.mobile.iscrollview.css" />
  <link rel="stylesheet" href="jquery.mobile/jquery.mobile.iscrollview-pull.css" />

  <script src="jquery.js"></script>
  <script type="text/javascript" src="jquery.mobile/jquery.mobile.js"></script>
  <script type="text/javascript" src="jquery.mobile/functions.js"></script>
  <script type="text/javascript" src="jquery.mobile/iscroll.js"></script>
  <script type="text/javascript" src="jquery.mobile/jquery.mobile.iscrollview.js"></script>
</head>
```

Fig 4 Appyling jQuery Mobile

Figure 4 shows the different codes provided by jQuery Mobile using HTML, CSS and Java script. Mobile web developers might want to change the theme, colors and other which is provided by jQuery Mobile platform and all you to do is copy the source provided.

5.1.2 Query Plan

This thesis is all about on how full text search optimization effectively helps in efficiently giving accurate results with faster retrieval time. There are two ways that it can optimize the search query, accuracy of results and decrease execution time. The general flow is shown on figure 2.

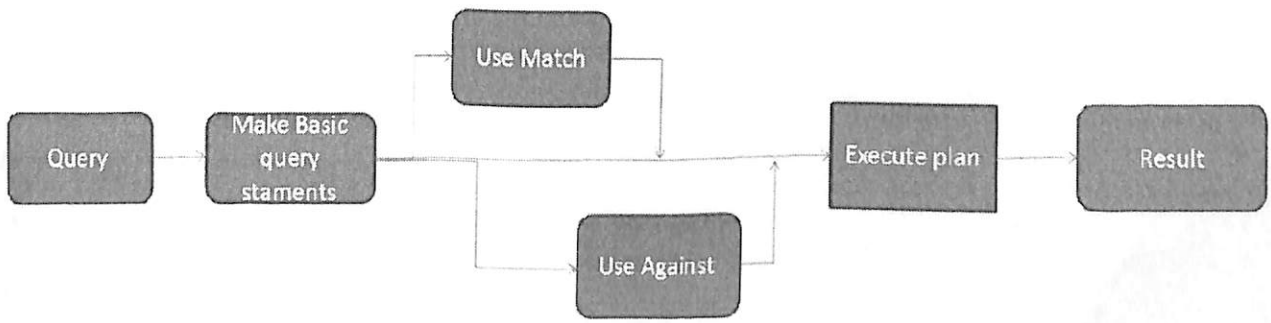
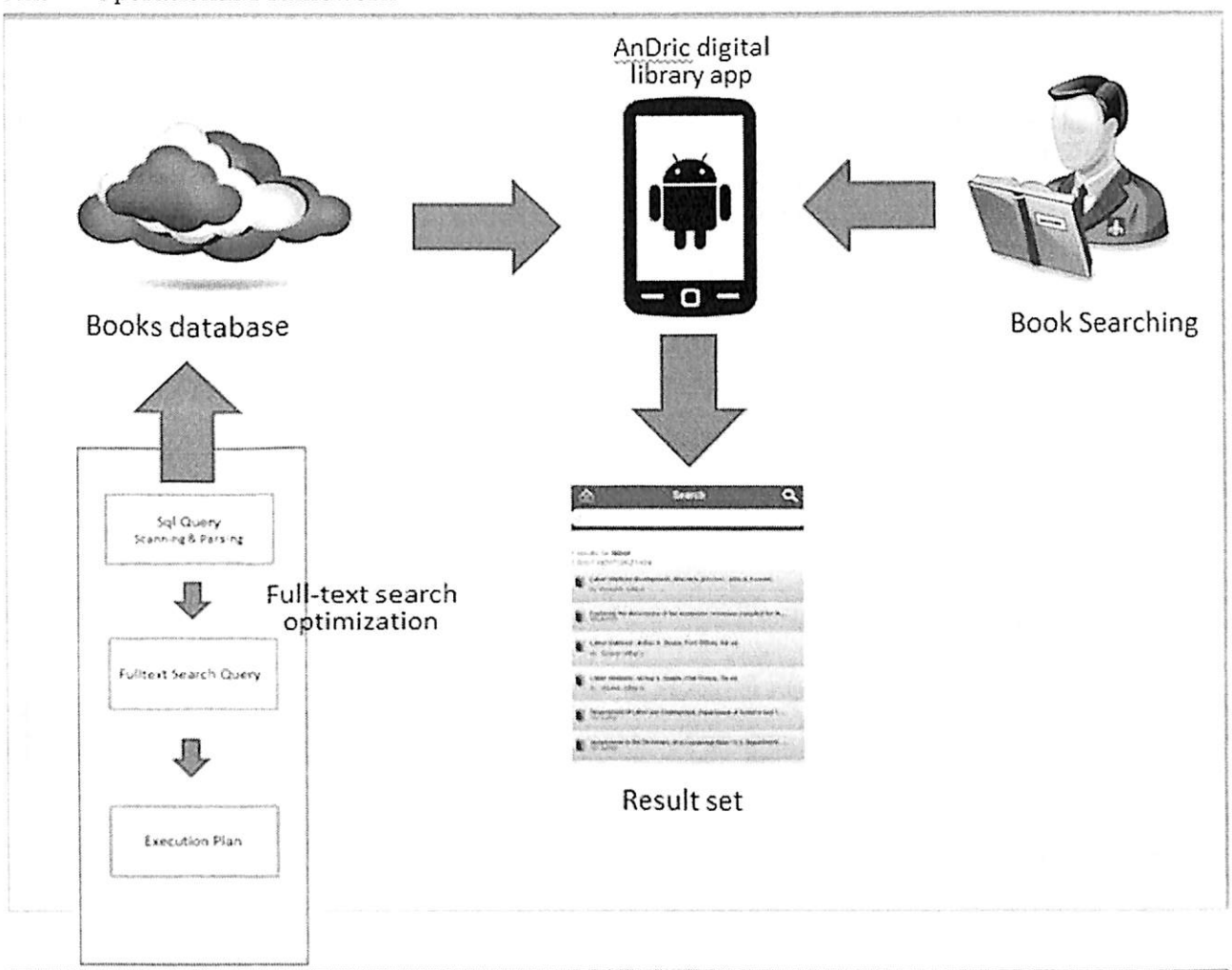


Fig 2 General flow of full text query optimization

- Make basic query statements: create a basic distribution plan that creates an individual pattern of queries.
- Optimization 1 "Use match": to improve the distribution plan that will give accurate results, by using match statement in the query.
- Optimization 2 "Use against": to improvement of distribution plan by decreasing the execution time.
- Execute plan: the last step where all search are made and retrieved results are joined together to create a whole result of the query with faster and accurate results.

5.1.3 Operational Framework



The implementation of the application starts with the display of the books available. Searching for books requires mobile data connection or Wi-Fi to connect to the server since it is an android application. Users can filter the search by choosing what specific type to search which can either be author, title of the book, or the book number. The main activity of the application starts when the user starts to search and display the results available. Full-text search optimization matches the word that is search which displays the results, shows number of books related to the searched word and the number of seconds that is retrieved from the database.

The app also displays the latest book that is available in the library.

5.2 IMPLEMENTATION

5.2.1 Creating Full-text Indexing and Searching

In performing the full text search, VARCHAR and TEXT columns that have been indexed with FULLTEXT can be used with special SQL statements. In order to use full text search, define the FULLTEXT index on the columns. Below is the example and the query statement that is used by the proponents.

ALTER TABLE anriclibrary ADD FULLTEXT(callNumber, author, title);

The proponents used in full-text search is a SELECT query with a MATCH...AGAINST clause. MATCH function is to specify the column names that have been identified in the FULLTEXT collection. The list of column inside the MATCH function must be exactly the same of the FULLTEXT index definition. AGAINST function takes a string to search for, and an optional modifier that indicates what type of search to perform. The query statement below is the query used by the proponents in searching books, author or call number.

SELECT book_id, title, author FROM book WHERE MATCH(callNumber, author, title) AGAINST('$search');

5.2.2 Integrating Full-text search optimization in Digital Library

```
<div data-role="page" >
<?php
error_reporting(E_ALL);
ini_set('display_errors', '1');
$author_id = "";
$book_id = "";
if(isset($_POST['search']) && $_POST['search'] != "" || isset($_GET['search']) && $_GET['search'] != "" ){
if(isset($_POST['search'])){
        $searchme = preg_replace("#[^a-z 0-9?/.']#i", "", $_POST['search']);
        $search = htmlspecialchars(stripslashes($searchme), ENT_QUOTES);
}else{
        $searchme = preg_replace("#[ a-z 0-9?/.']#i", "", $_GET['search']);
        $search = htmlspecialchars(stripslashes($searchme), ENT_QUOTES);
}

    $time_start = microtime(true);
    $sqlCommand = "SELECT book_id FROM book WHERE MATCH(callNumber, author, title) AGAINST '$search'i";


    $query = mysql_query($sqlCommand) or die(mysql_error());
    $count = mysql_num_rows($query);
    $time_end = microtime(true);
    $time = $time_end - $time_start;
```

Figure 1

In order for the search function to work, search function of the library is integrated in Full-text search optimization in order to connect to database. Figure above shows the search function and the full-text search optimization combined.

5.2.3 Deployment of Digital library in Android platform

package com.thesis.andriclibrary;

import android.os.Bundle;

```
import android.annotation.SuppressLint;
import android.app.Activity;
import android.view.Menu;
import android.webkit.WebSettings;
import android.webkit.WebView;

@SuppressLint("SetJavaScriptEnabled") public class MainActivity extends Activity {

    @Override
    protected void onCreate(Bundle savedInstanceState) {
        super.onCreate(savedInstanceState);
        setContentView(R.layout.activity_main);
        WebView myWebView = (WebView) findViewById(R.id.webview);
        WebSettings webSettings = myWebView.getSettings();
        webSettings.setJavaScriptEnabled(true);
        myWebView.loadUrl("http://192.168.43.71/thesis/");
    }

    @Override
    public boolean onCreateOptionsMenu(Menu menu) {
        // Inflate the menu; this adds items to the action bar if it is present.
        getMenuInflater().inflate(R.menu.main, menu);
        return true;
    }
```

After developing the library application in HTML using JQuery mobile and integrating search function to Full-text search optimization, the proponents deployed the application into android platform for it to be usable in smartphones. The code above shows the deployment of the application using java.

5.3 EVALUATION

5.3.2 Usability of the application

The proponents went to Ateneo de Davao's University library to conduct a survey on the application. To rate the research project in its usability by letting the students try the application. Ranges of students who tried the application is from 17-21 years of age and 20 of the students tested the application. Survey sheets were given to them and answered 12 questions regarding on the usability of the application.