

# **CAT 400 Undergraduate Major Project**

## **Analysis Review Presentation**

*supervisor*

**Dr. Gan Keng Hoon**

*examiners*

**Associate Professor Dr. Azman Bin Samsudin,**

**Professor Dr. Mandava Rajeswari**

*prepared by*

***Gan Kian Min***

***110938***



# Expert Search - The Claw

*Project ID*

**SC141521**

*Category*

**Intelligent System**

# Introduction

# Project Background

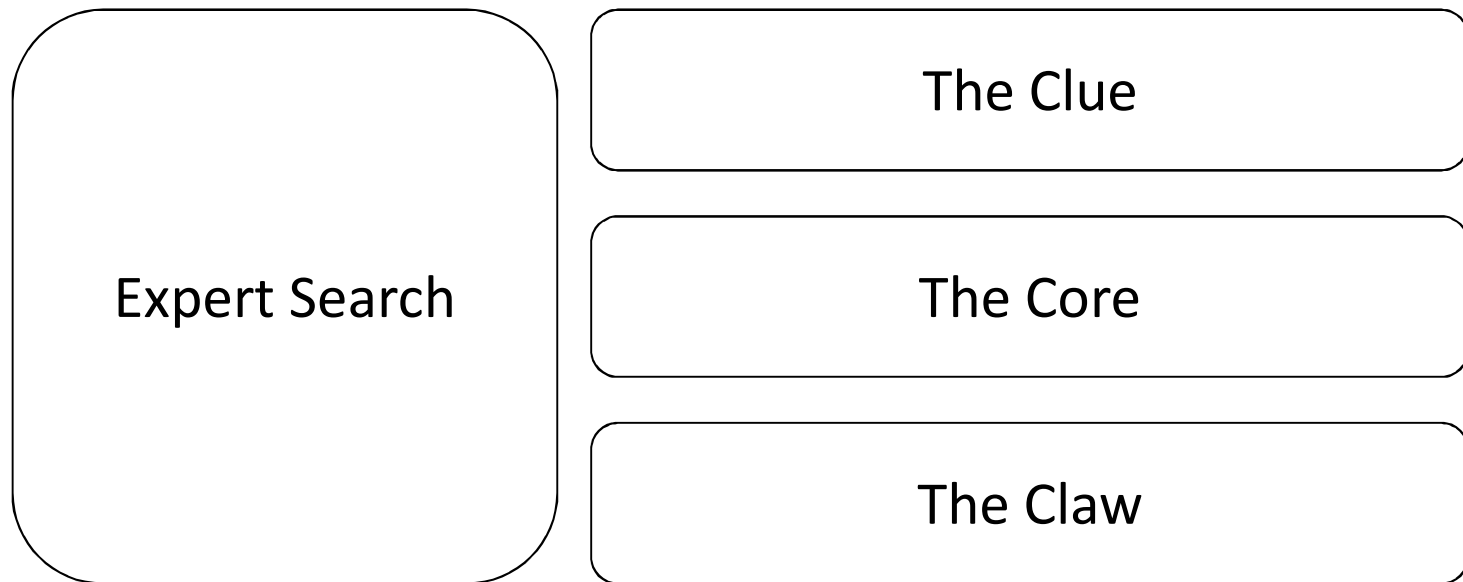
Search Engine System

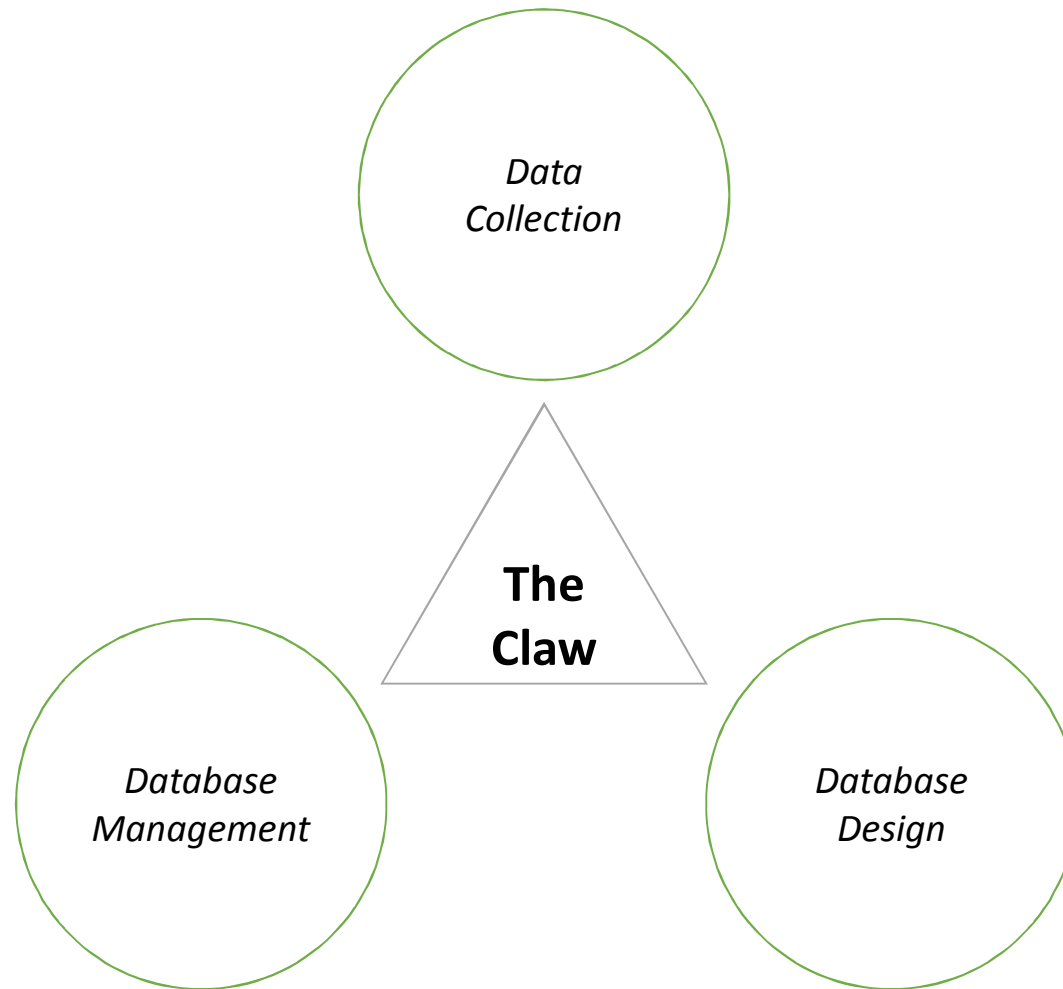
Focus on CS USM Academic Staffs

Academic Information

Research Field

# Modules

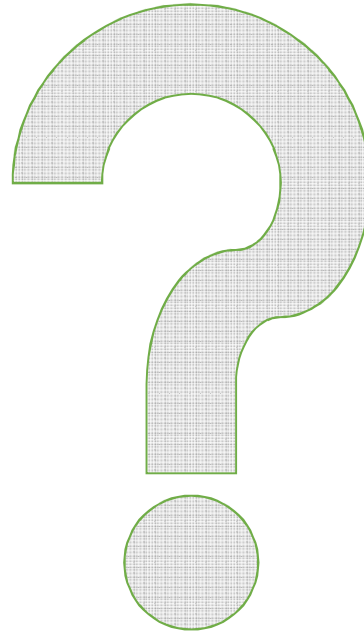




# Problem Statement

## Data Collection Process

- Manual work
- Time Consume
- Need Automate System
- Data Error in automate process



## Database Design & Management

- Need an Manual “Environment” to control data
- Admin System
- Manage database

# Proposed Solutions

**Automate Data  
Extraction Tools**

**Database Management  
Admin Tools**



# Data Sources

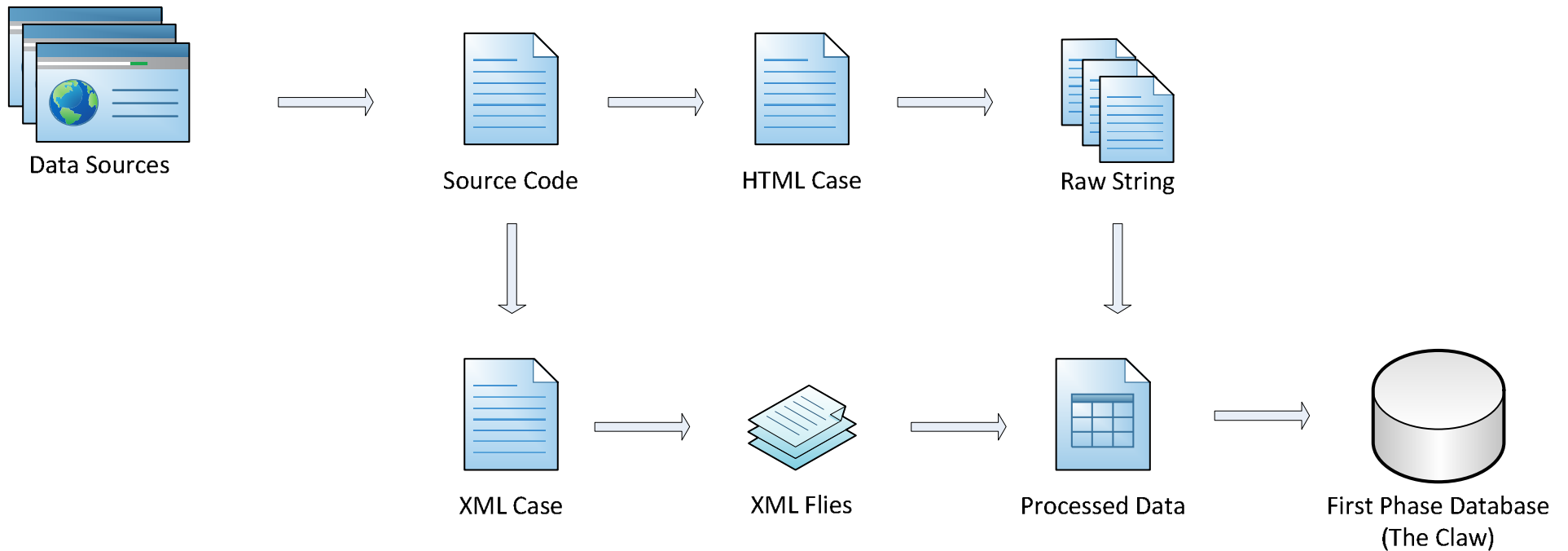


DBLP

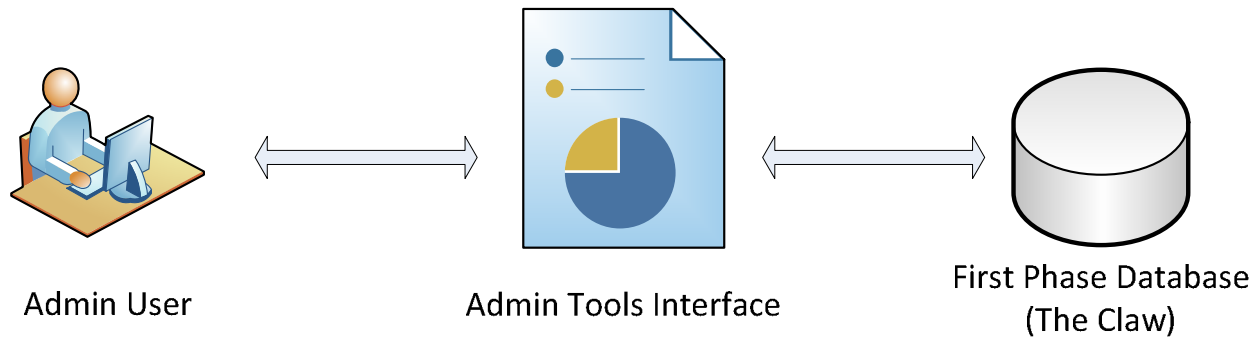


CS USM Staff Website

# Data Extraction Tools



# Database Management Admin Tools



- *Data Table Form*
- *Bulk insert, update, delete*
- *XML Generator*

# System Objective



**The Claw  
Of  
Expert Search**

Identify Suitable Information

Design Expert Search Database

Develop Automate Data  
Extraction Tools

Develop Database Management  
Tools

# Benefits & Uniqueness

## Benefits

Reduce Data Collection  
Workload

Ease Maintain Database  
System

Better environment for  
future development

## Uniqueness

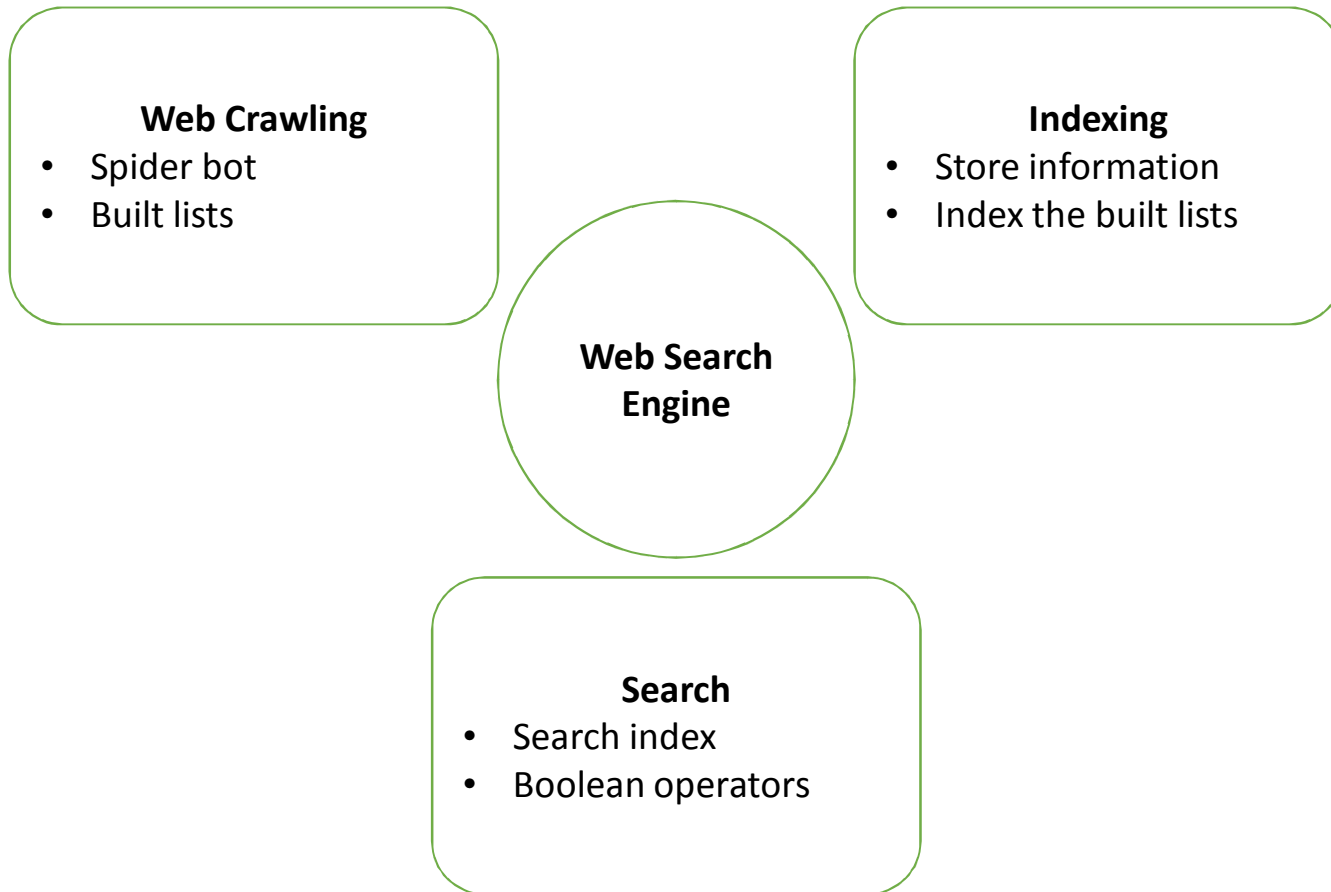
First Search Engine System  
Focus on USM CS Staff

Up-to-date Bibliography  
Information

Academic Information

# Related Work

# Related Technology



# Existing Systems



# Google Scholar

Web Images More...

Sign in

 My library  My Citations  Alerts  Metrics  Settings



Articles  include patents  Case law

**Stand on the shoulders of giants**

**Search all scholarly literature from one convenient place**

**Explore related works, citations, authors, and publications**

**Locate the complete document through your library or on the web**

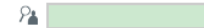
**Keep up with recent developments in any area of research**

**Check who's citing your publications, create a public author profile**

# DataBase systems and Logic Programming (DBLP)



[home](#) | [browse](#) | [search](#) | [about](#)



## The DBLP Computer Science Bibliography

maintained by the DBLP Team

Welcome to DBLP. This server provides bibliographic information on major **computer science journals and proceedings**. DBLP indexes more than 2.6 million articles and contains many links to home pages of computer scientists.

Search for a **person**:    ->Help

General search on DBLP: **CompleteSearch** - **Faceted search** - **Free search**

**Conferences & Workshops:** A - B - C - D - E - F - G - H - I - J - K - L - M - N - O - P - Q - R - S - T - U - V - W - X - Y - Z

**Journals:** A - B - C - D - E - F - G - H - I - J - K - L - M - N - O - P - Q - R - S - T - U - V - W - X - Y - Z

DBLP is available from three hosts: Trier I - Trier II - Dagstuhl

**Your Contribution ...** : Please read the **FAQ Pages**. Send an e-mail to [ley@uni-trier.de](mailto:ley@uni-trier.de), if ...

- ... you notice errors or omissions. I try to unify different spellings of names, often this isn't possible without your help.
- ... you are listed, but a link to your home page is missing or broken.
- ... if you edit a journal or proceedings and want this publication to be included into the DBLP server.

Disclaimer: All information on the DBLP server is provided in good faith, we take no responsibility for the accuracy of the information. Please read the copyright page.

**Bibliographic information on major computer science publications**

**Mission of *dblp* is to support computer science researchers**

**Providing free access to high-quality bibliographic meta-data and links to the electronic editions of publications**

# Stanford xSearch

Select a subject category, a document type, or both to find types within a category.  
Can search multiple categories and types. Click on a subcategory to select individual sources.

**Keyword:**

**Title:**

**Author:**

**Match:**  **Field(s)**

**Date Range:**  **to**

## Categories

- Business**
- Engineering**
- Humanities**
- International & Area Studies**
- Law**
- Medicine**
- Sciences**
- Social Sciences**

## Document Types

- [Archival Finding Aids](#)
- [Data & Statistics](#)
- [Databases](#)
- [Full-text Books](#)
- [Full-text Journals](#)
- [Grants & Funding](#)
- [Government Documents](#)
- [Images](#)
- [Streaming Media](#)

**Provides stanford students and researchers with a single search option for multiple online resources.**

**For the initial launch, 28 article databases and ejournal/ebook platforms were available.**

**xSearch used to include 170 and then was expanded to 200 resources that include a broad array of subject areas and types of materials.**

# System Requirement & Analysis

# Project Scope

## The Claw

### Database Design

Determine Purpose

Organize information

Design tables

Setup tables relationship

Apply Normalization Rule

### Database Management

Interface Design

Features Design

Table Display, Editable from

Insert, Update, Delete, Search

### Data Collection

HTML Automate data extraction tool

XML Automate data extraction tool

Database bulk insert tool

XML Generator tool



# Capabilities

Back-End System

Automate Data Extraction Tools

Database Management Admin Tool

XML Generator

# Limitations

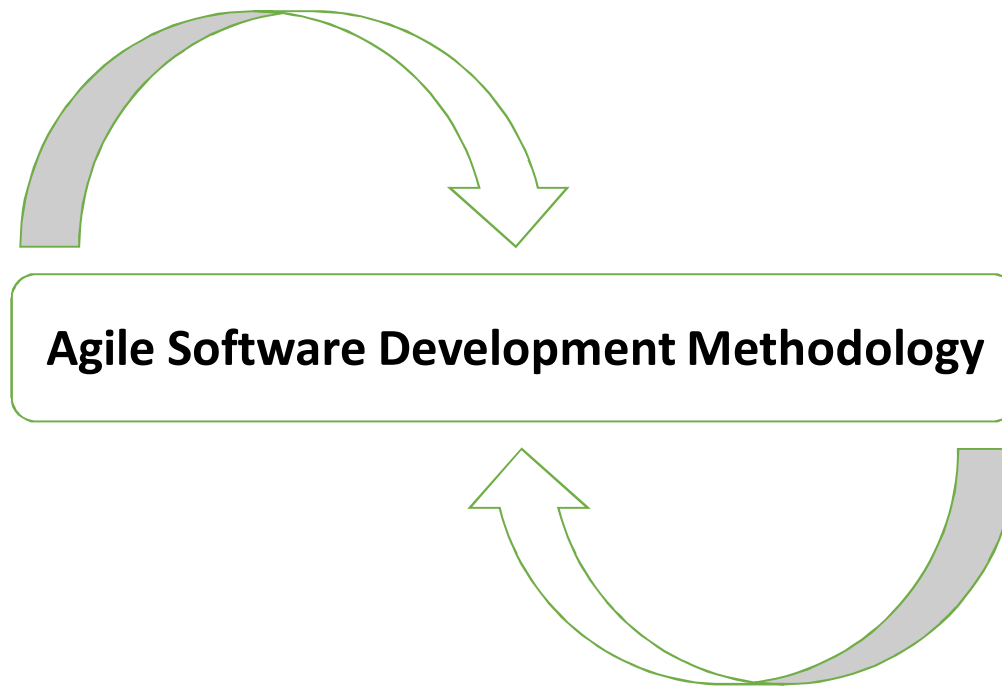
Data Sources

One tool for one source

Automate Data Extraction Tools

Change of Coding & Algorithm

# Methodology



# Cases, Problems and Solutions

# Case I



Bibliography

Same Author, Different Citation



Group Different Citations

Identify as Same Person




**dblp**  
computer science bibliography

[home](#) | [browse](#) | [search](#) | [about](#)

 Gan Keng Hoon 

## Search results for "Gan Keng Hoon"

> [Home](#)

 [Trier 2](#)

### Author results

#### Exact matches

- [Gan Keng Hoon](#)

#### Likely matches

- [Gan Keng Hoon](#)
- [Keng Hoon Gan](#)

# Algorithms

## The Levenshtein Distance

Spell Check

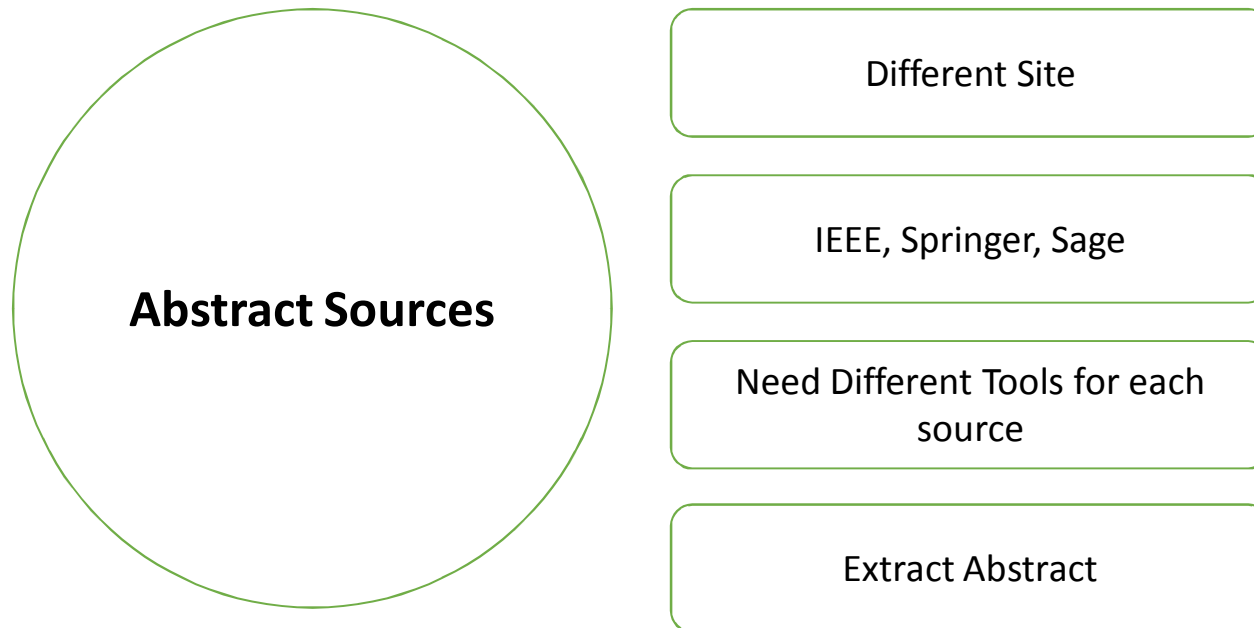
Replace  
Corrected  
String

Compare  
TWO  
Strings

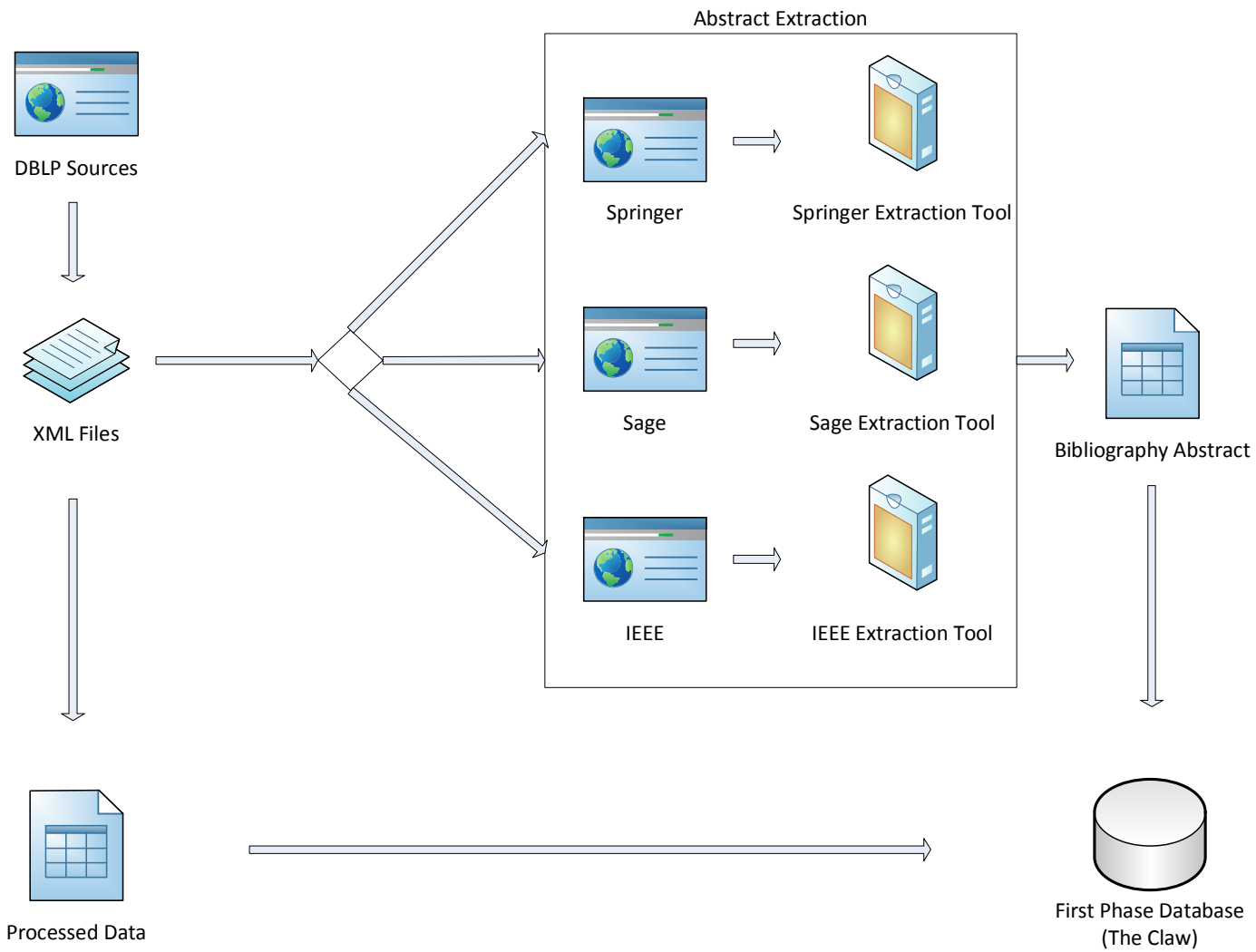
Find  
Shortest  
Distance

Identify  
Similar  
Text

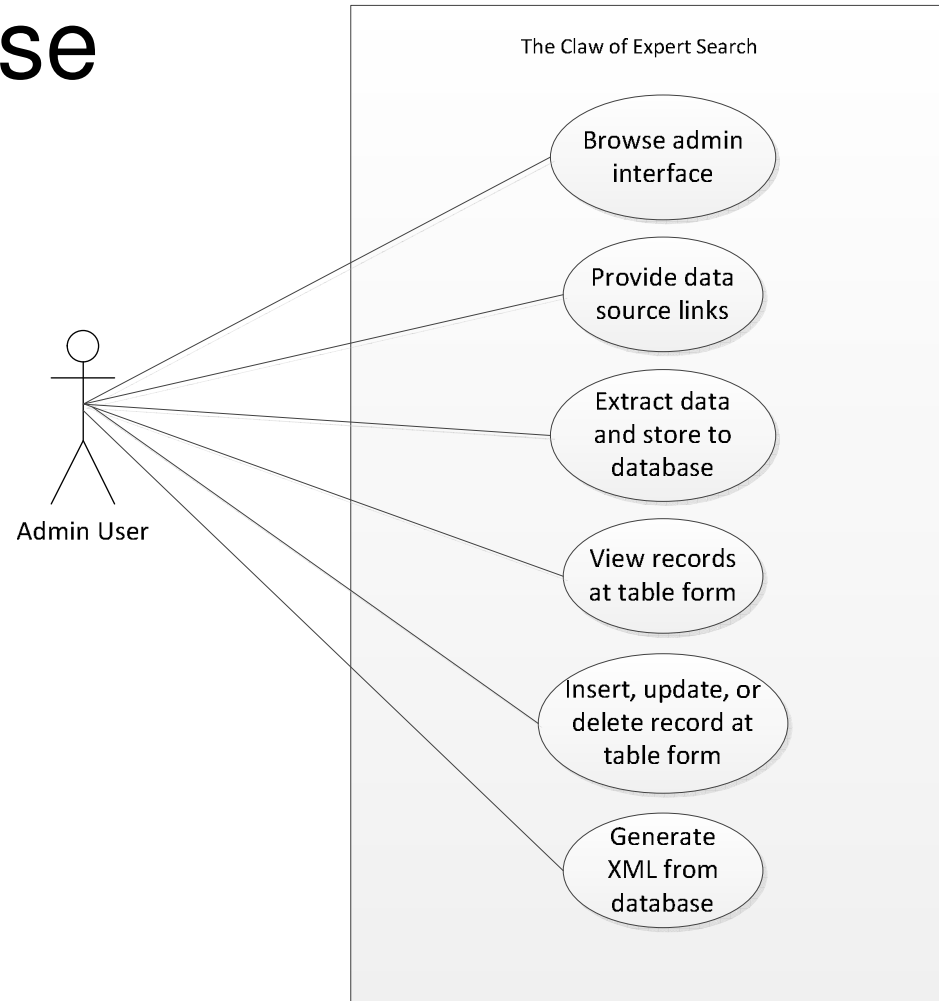
# Case II



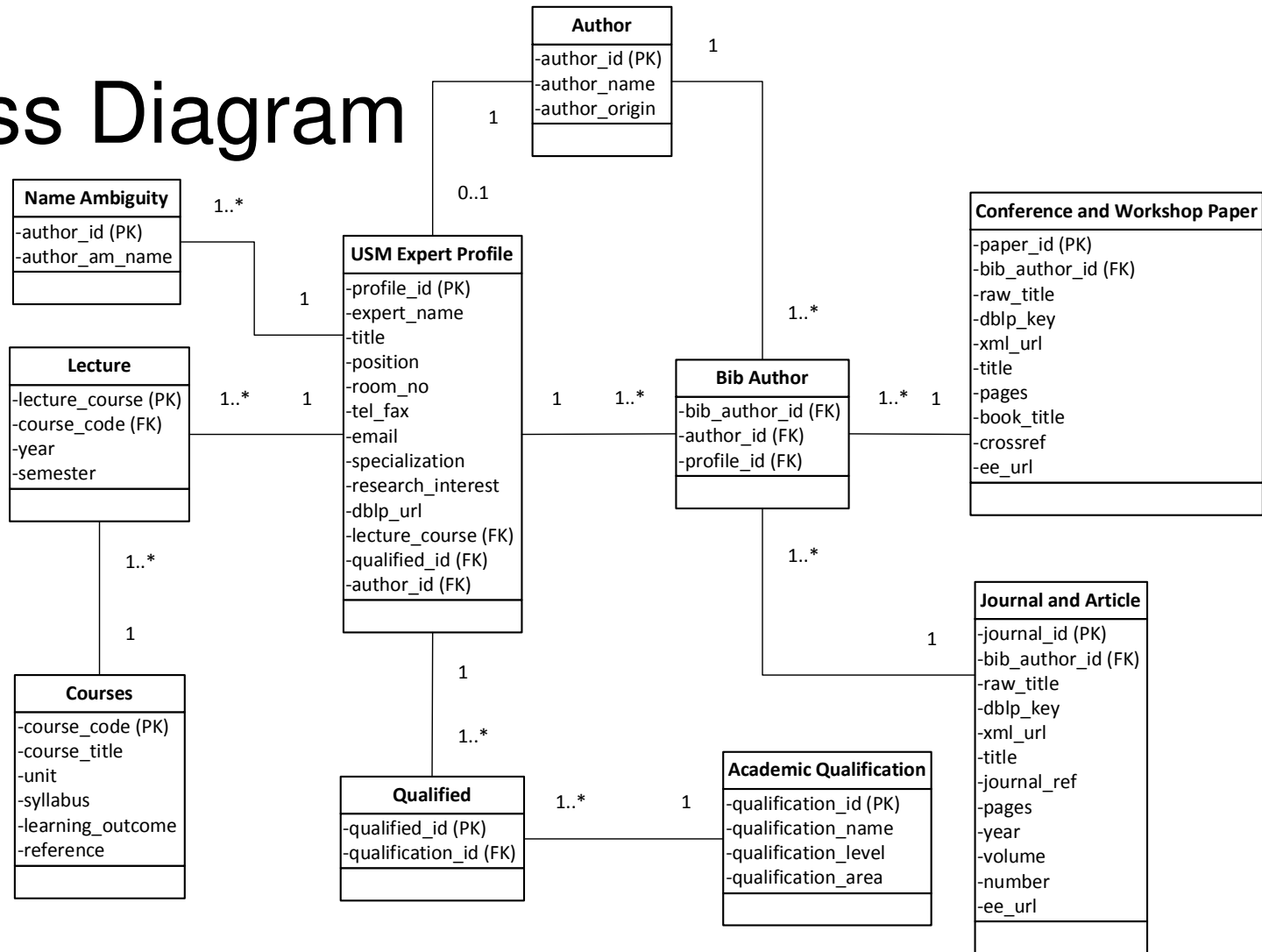




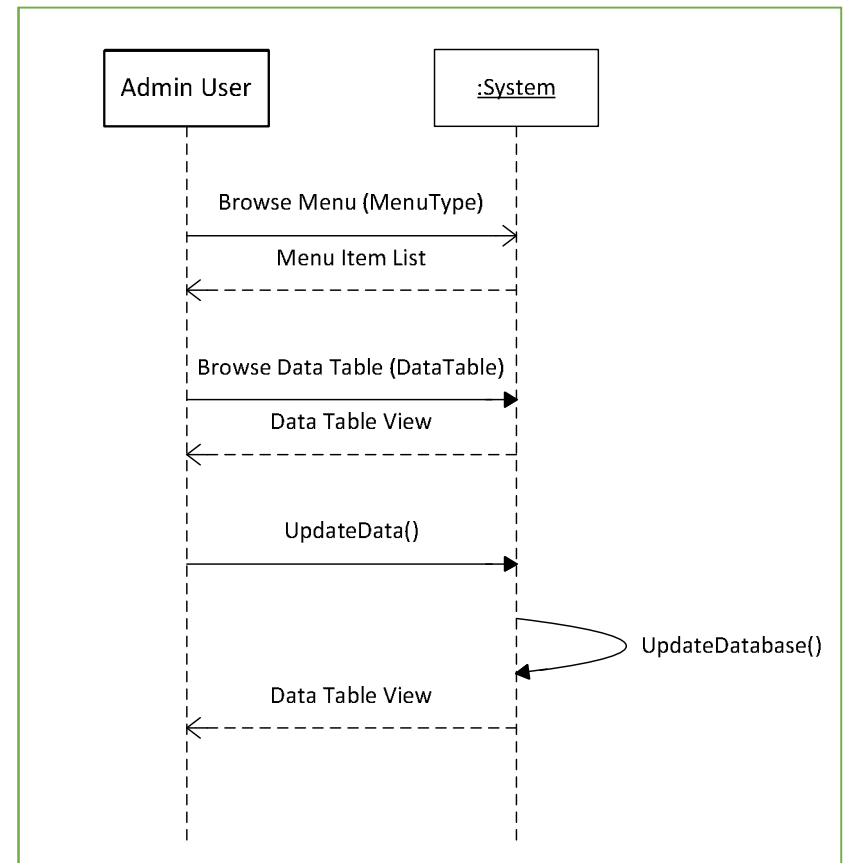
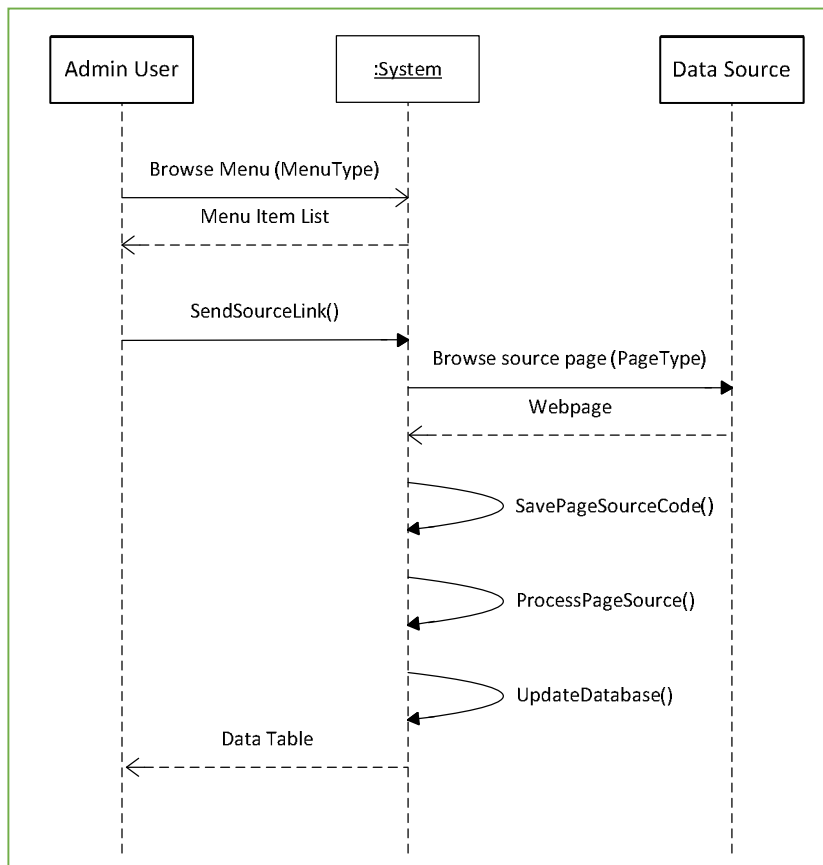
# Use Case



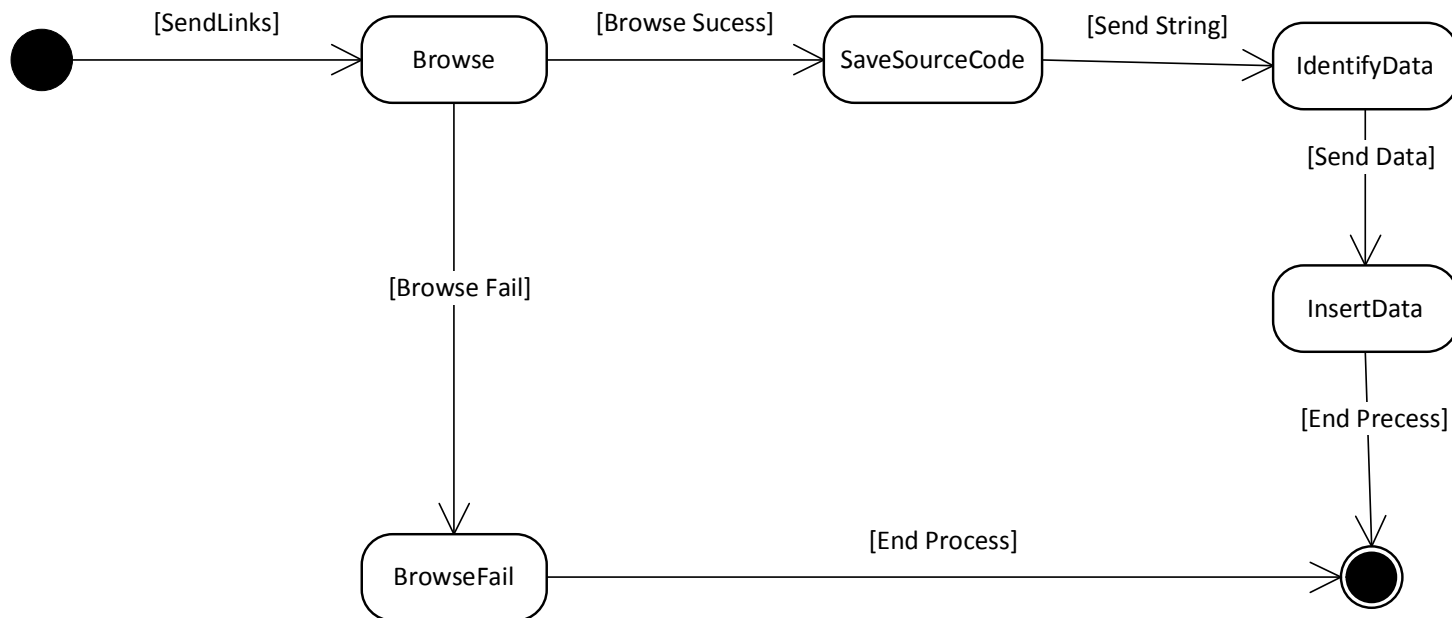
# Class Diagram



# Sequence Diagram



# State Diagram



**Conclusion**

Admin User

Manage Back-End Tasks

Prepared Database for The Core Module

Future Development

