بسم الله الرحمن الرحيم

Sudan University of Science & Technology
College of Graduate Studies

Search Engines With Particular Reference To Multilingual Search

Supervised By:

Prof. Izzeldin Mohammed Osman

Prepared By:

Wegdan A. Elhameed

June 2005

Abstract

- Search Engines are a useful tool to retrieve information from internet.
- Although, internet users speak different languages most of resources are written and published in English.
- This research investigates multilingual search, to help people finding useful contents stored in multiple languages and examines the case of English and Arabic languages.
- The model presented for multilingual search engine is illustrated by a practical example of Sudan University of Science and Technology journal site.

Chapter 1: Literature Review

- Introduction
- DNS (Domain Name Server)
- Web Servers
- Meta Tags
- User Requirements Of The Search Engine
- Search Engine Specification
- Needs To Multilingual Search
- Research Problem
- Research Overview
- Research Goals And Objectives
- Related Works

Chapter 2: Internet Search Engines

- Definitions
- History
- How they work?
- Components
- Query Processing (Query Engine)
- Optimizing Query Execution and Information Retrieval
- Features and Characteristics
- Search Engine Architecture
- Challenges
- Problems
- Spamming and Cloaking

Chapter 3: Google Search Engine

- Design Goals
- System Features
- Anchor Text
- Differences Between the Web and Well Controlled Collections
- System Anatomy
- Crawling the Web
- Searching
- Performance

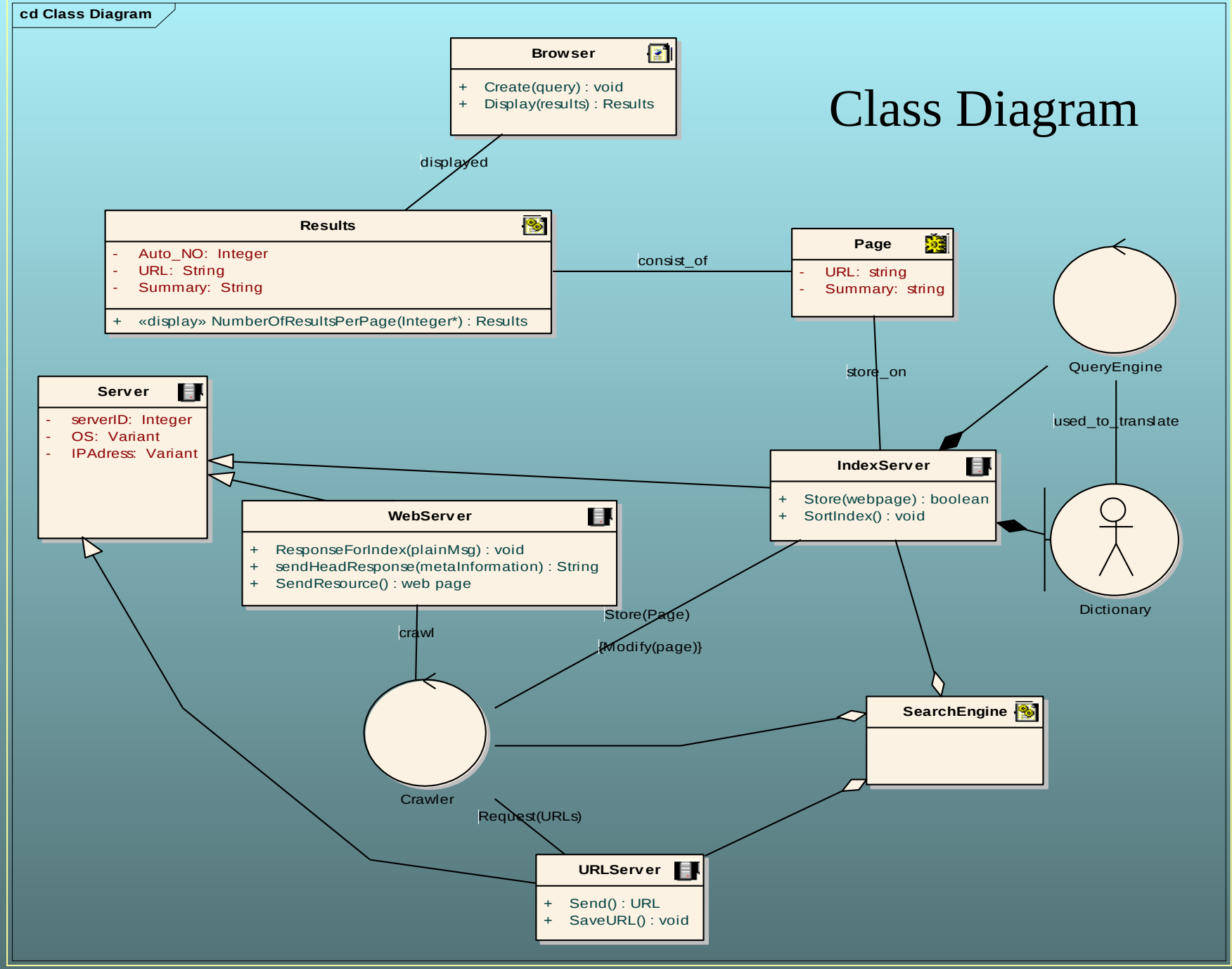
Chapter 4: Intranet Search Engines

- Definition
- Challenges
- How it Work?
- Gathering Documents and Indexing
- Intranet Metadata
- Intranet Search
- Search Problems
- How to make a Good Intranet Search Engine?
- Intranets vs. Internet : Axioms
- Intranets vs. e Internet: Structural Differences

Chapter 5: Multilingual Search Engine Model

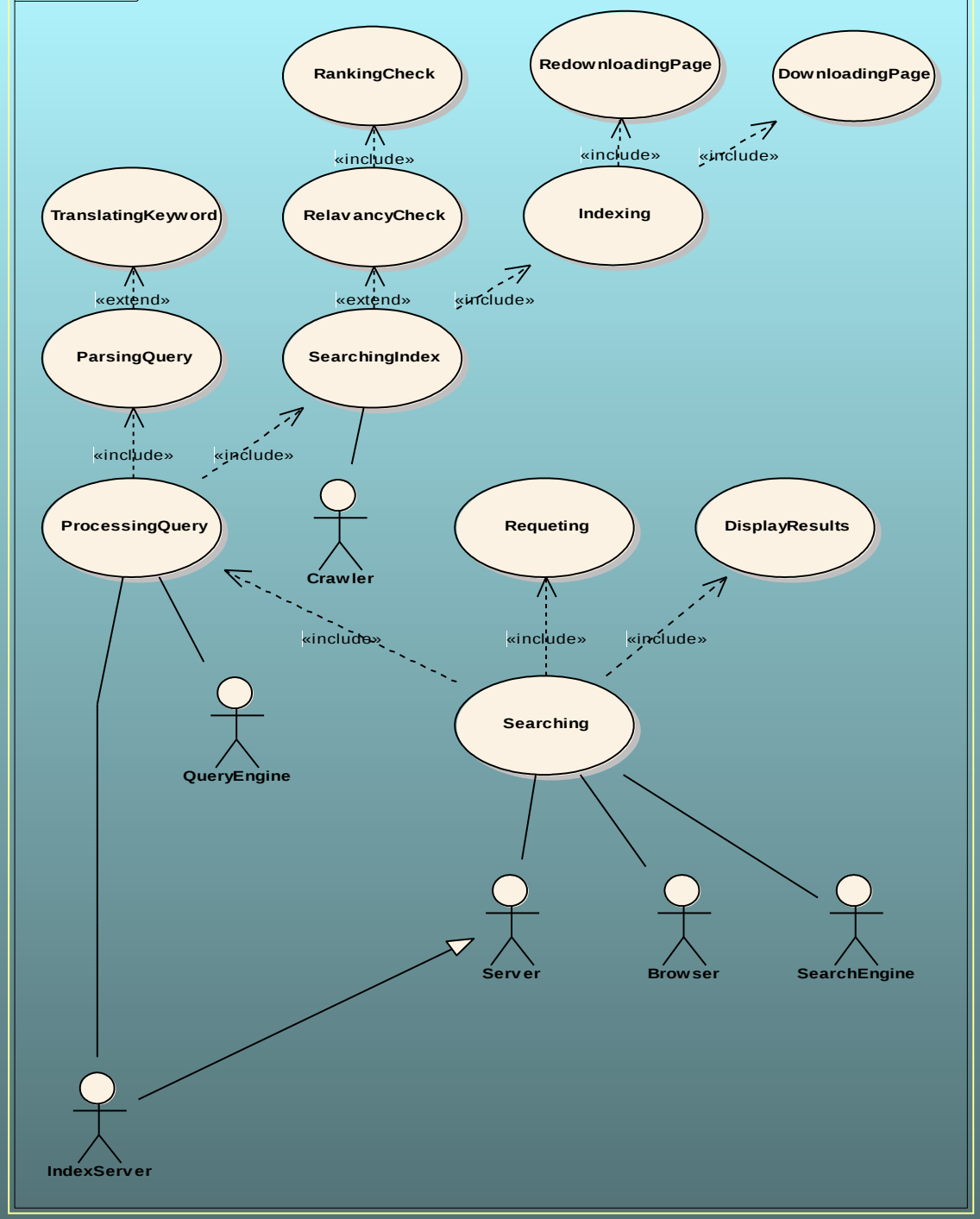
- Logical Model
  - Class Diagram
- Use Case Model
  - Use Cases Views
- Dynamic Model
  - Activity Diagram
  - Sequence Diagram

Logical Model

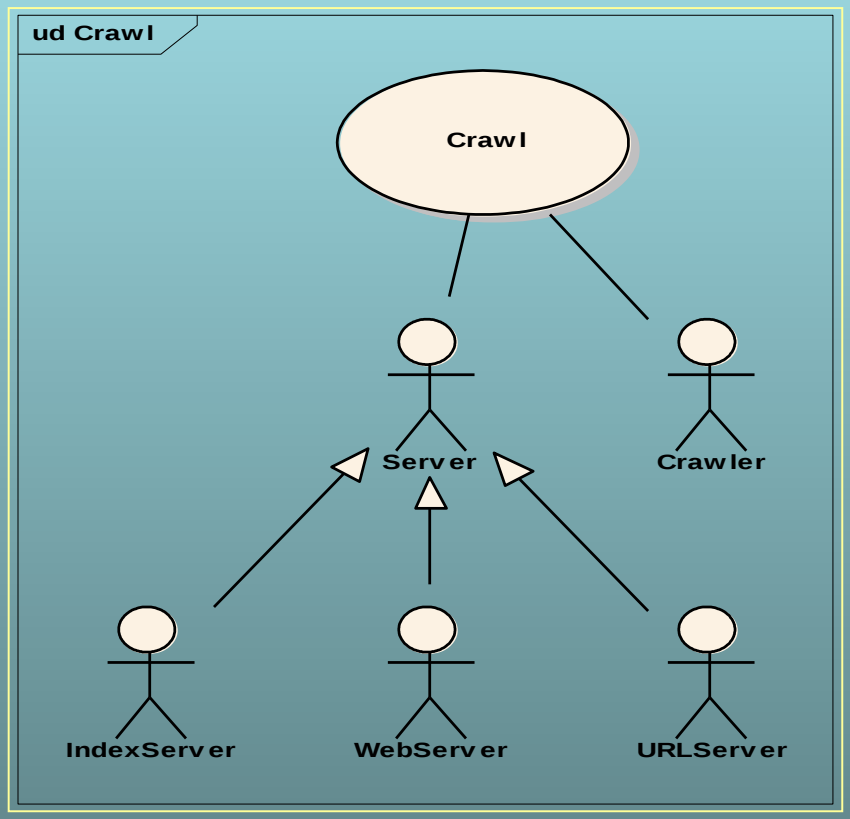

cd Class Diagram
Class Diagram
Browser
+ Create(query) : void
+ Display(results) : Results
displayed
Results
- Auto_NO: Integer
- URL: String
- Summary: String
+ «display» NumberOfResultsPerPage(Integer*) : Results
consist_of
Page
- URL: string
- Summary: string
QueryEngine
store_on
used_to_translate
Server
- serverID: Integer
- OS: Variant
- IPAdress: Variant
IndexServer
+ Store(webpage) : boolean
+ SortIndex() : void
WebServer
+ ResponseForIndex(plainMsg) : void
+ sendHeadResponse(metaInformation) : String
+ SendResource() : web page
Dictionary
Store(Page)
{Modify(page)}
crawl
SearchEngine
Crawler
Request(URLs)
URLServer
+ Send() : URL
+ SaveURL() : void

Use Case Model

Searching Use Case

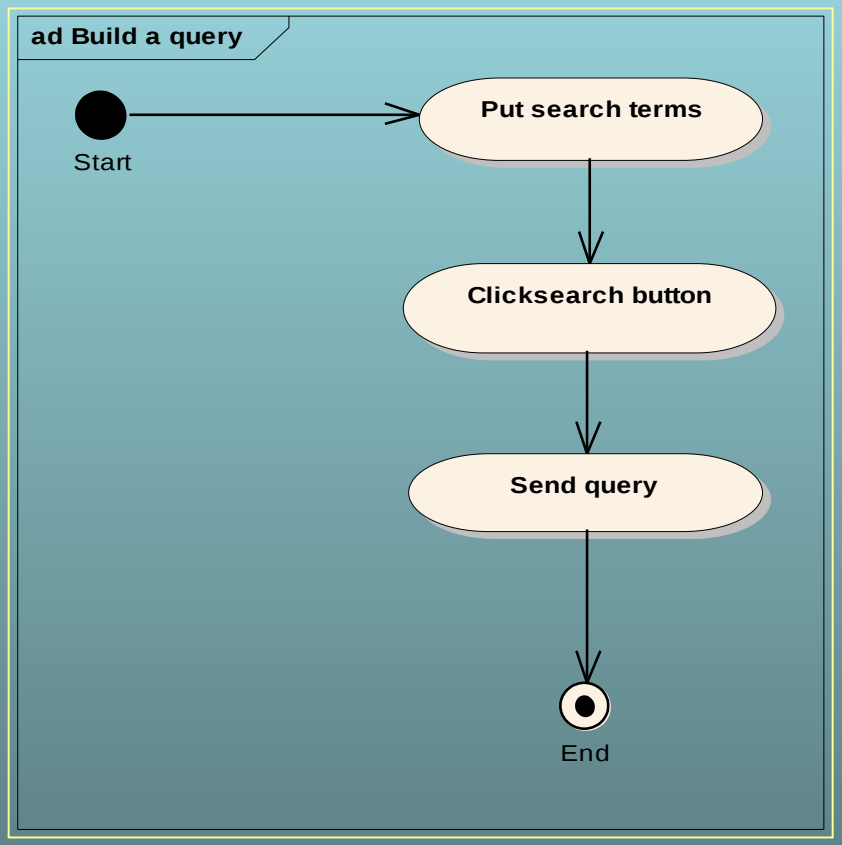


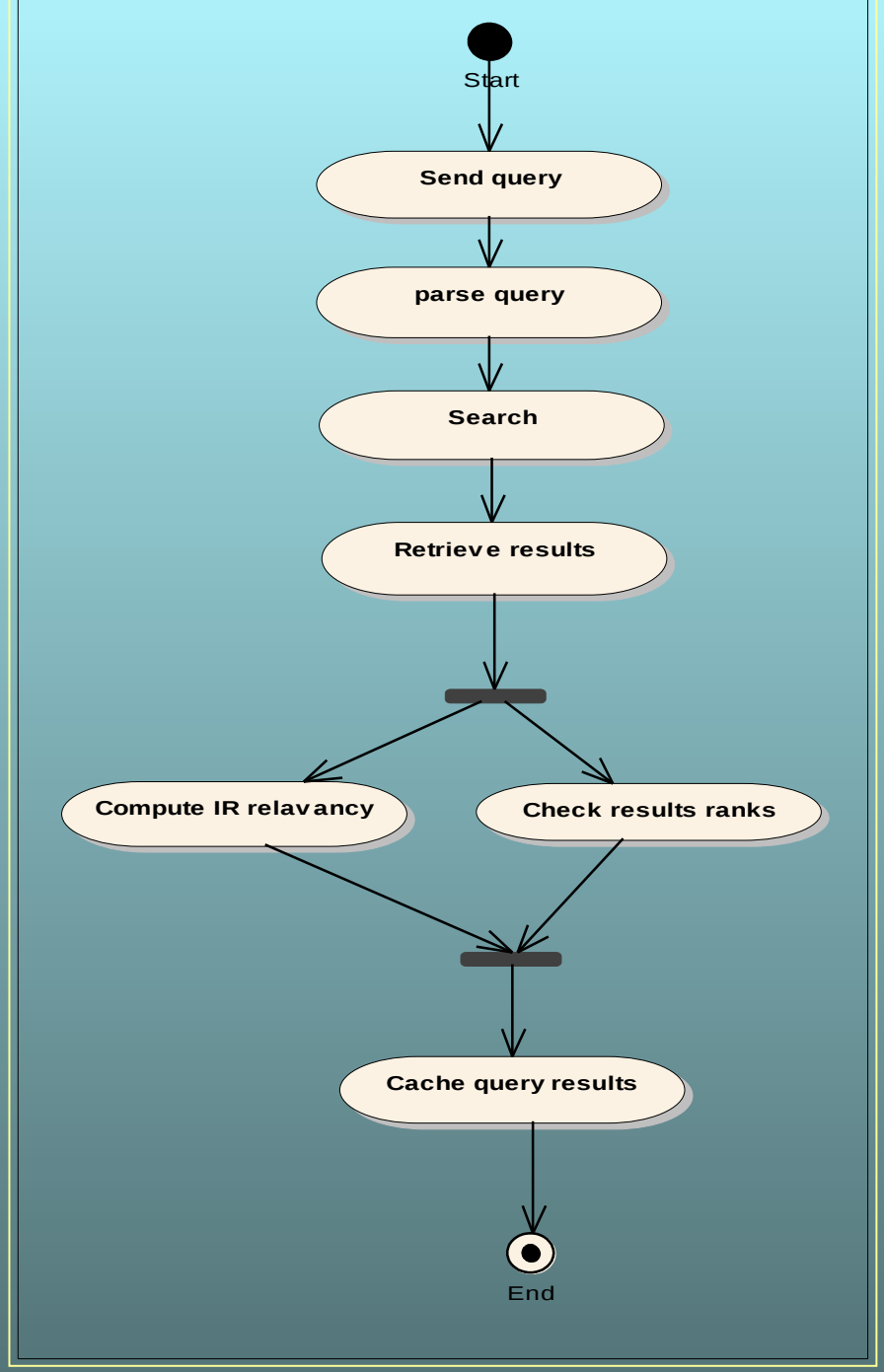
Crawl Use Case



Dynamic Model

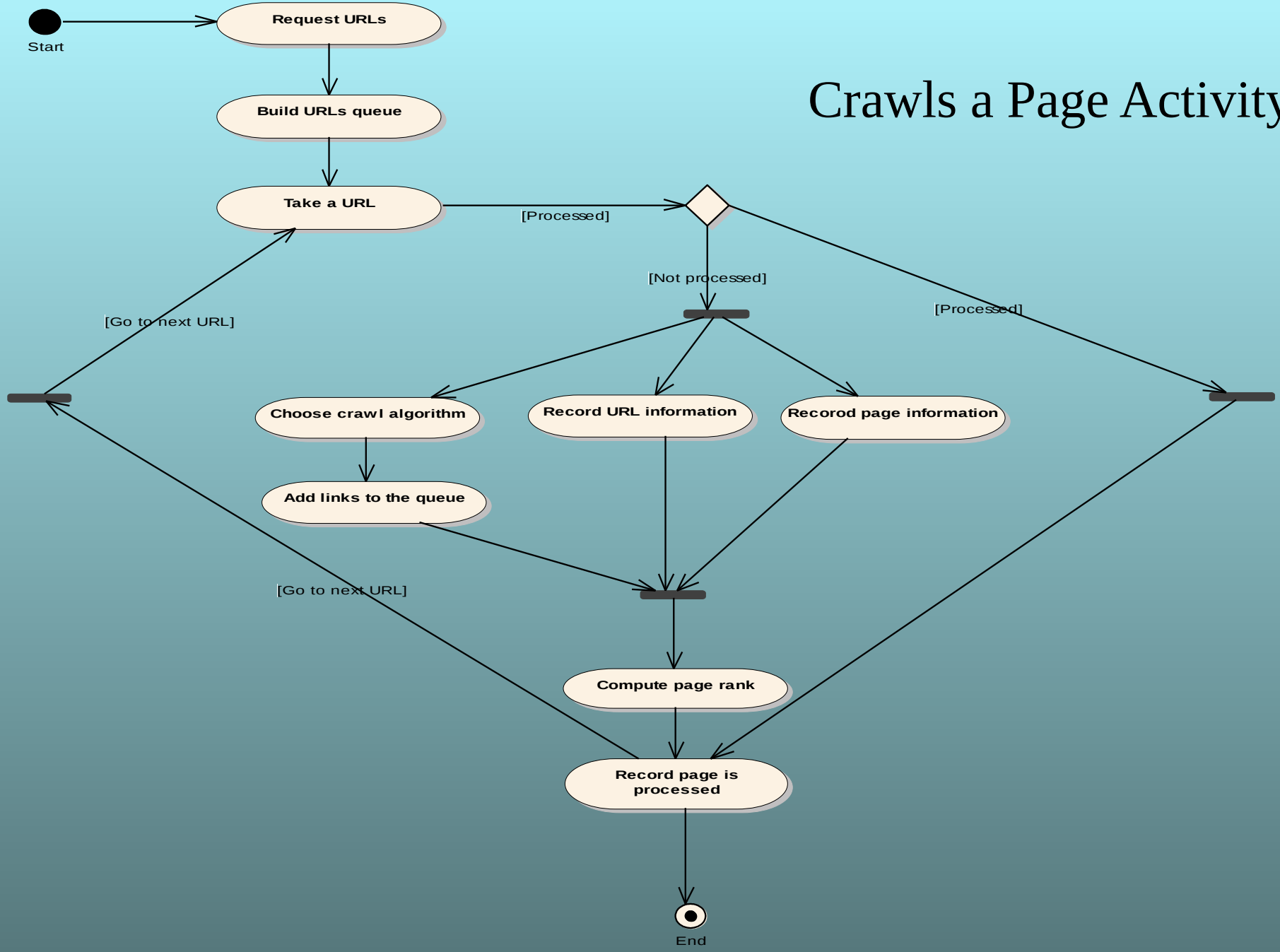
Build a Query Activity



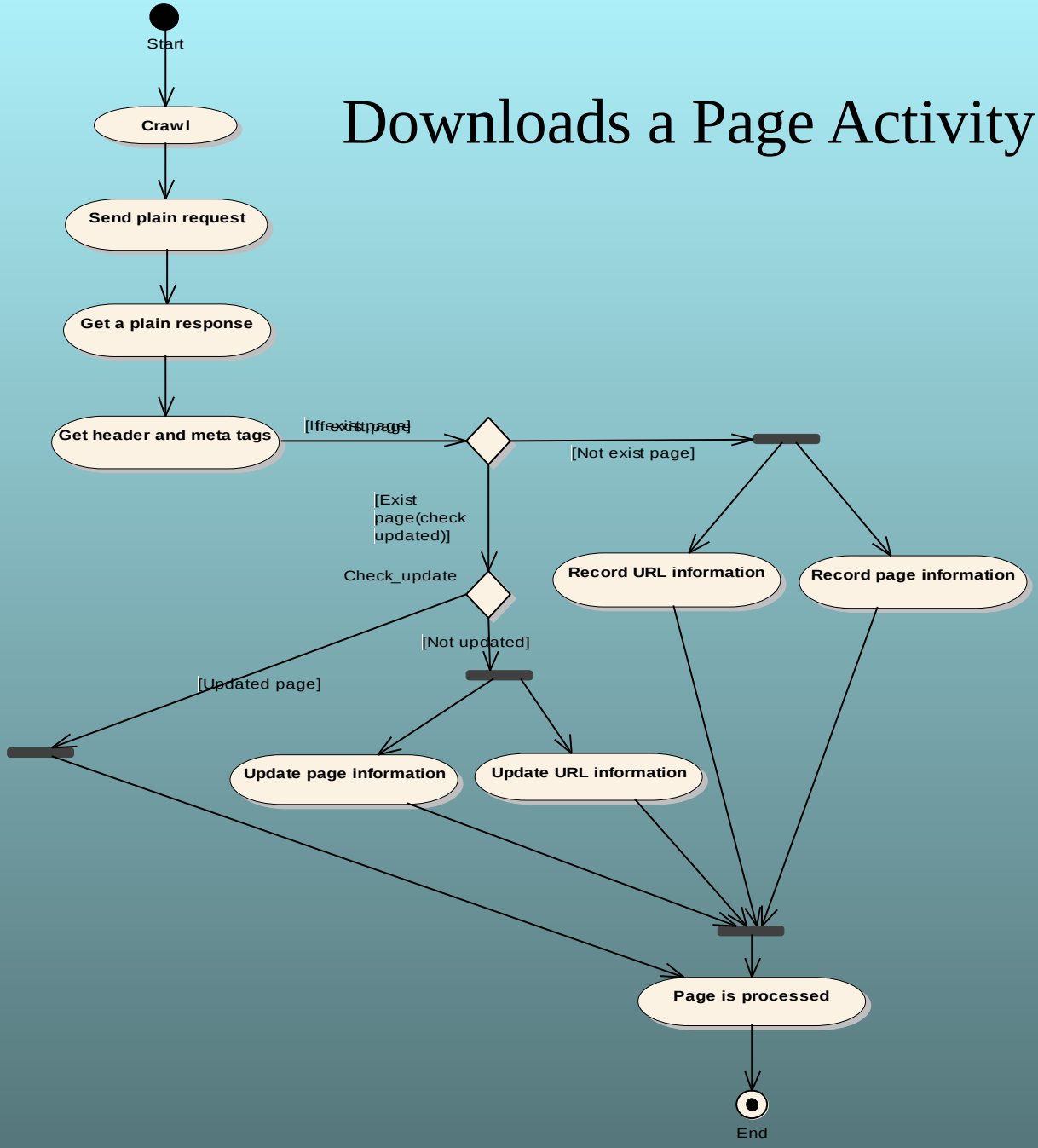


Check Relevancy Activity

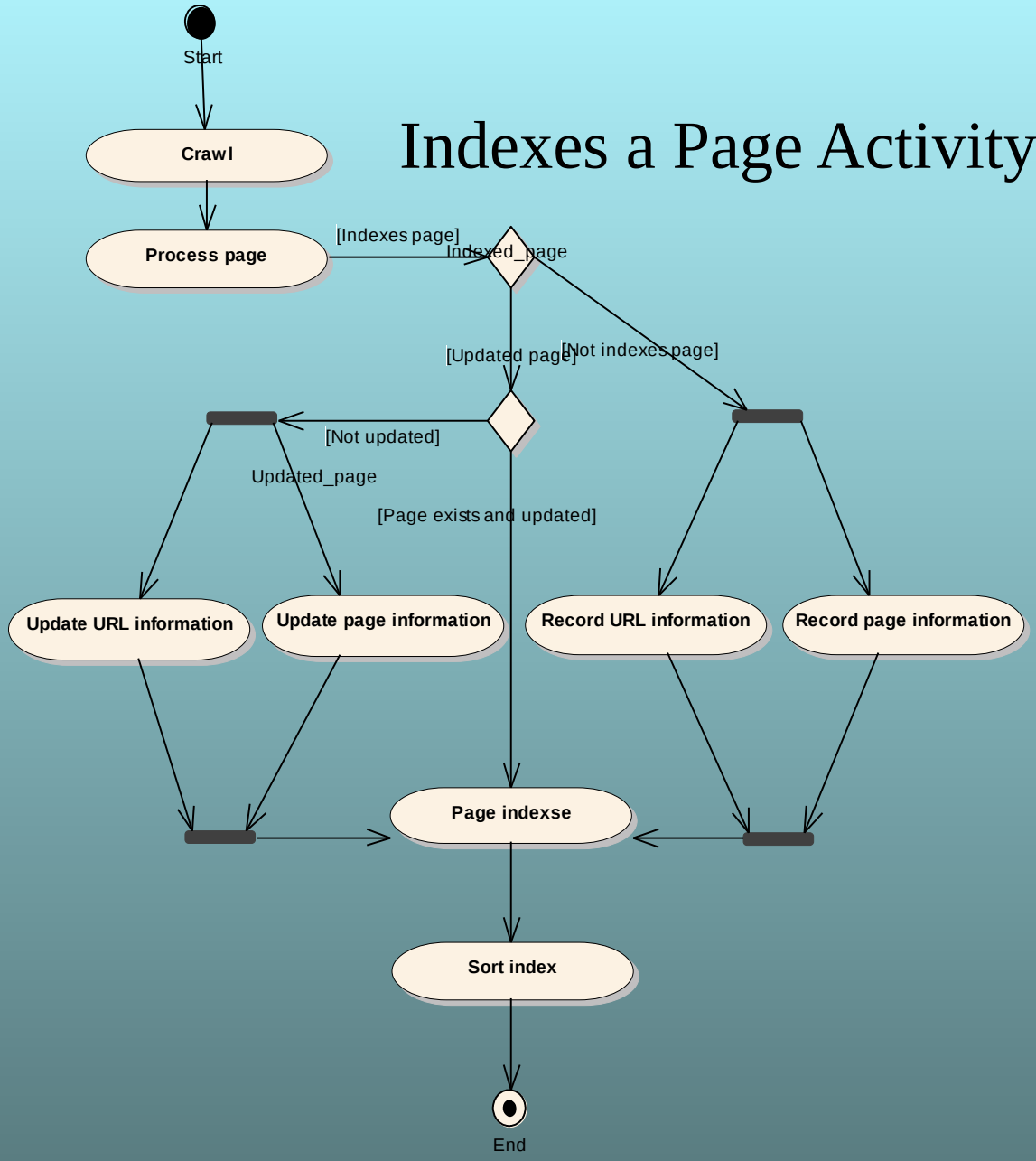
Crawls a Page Activity

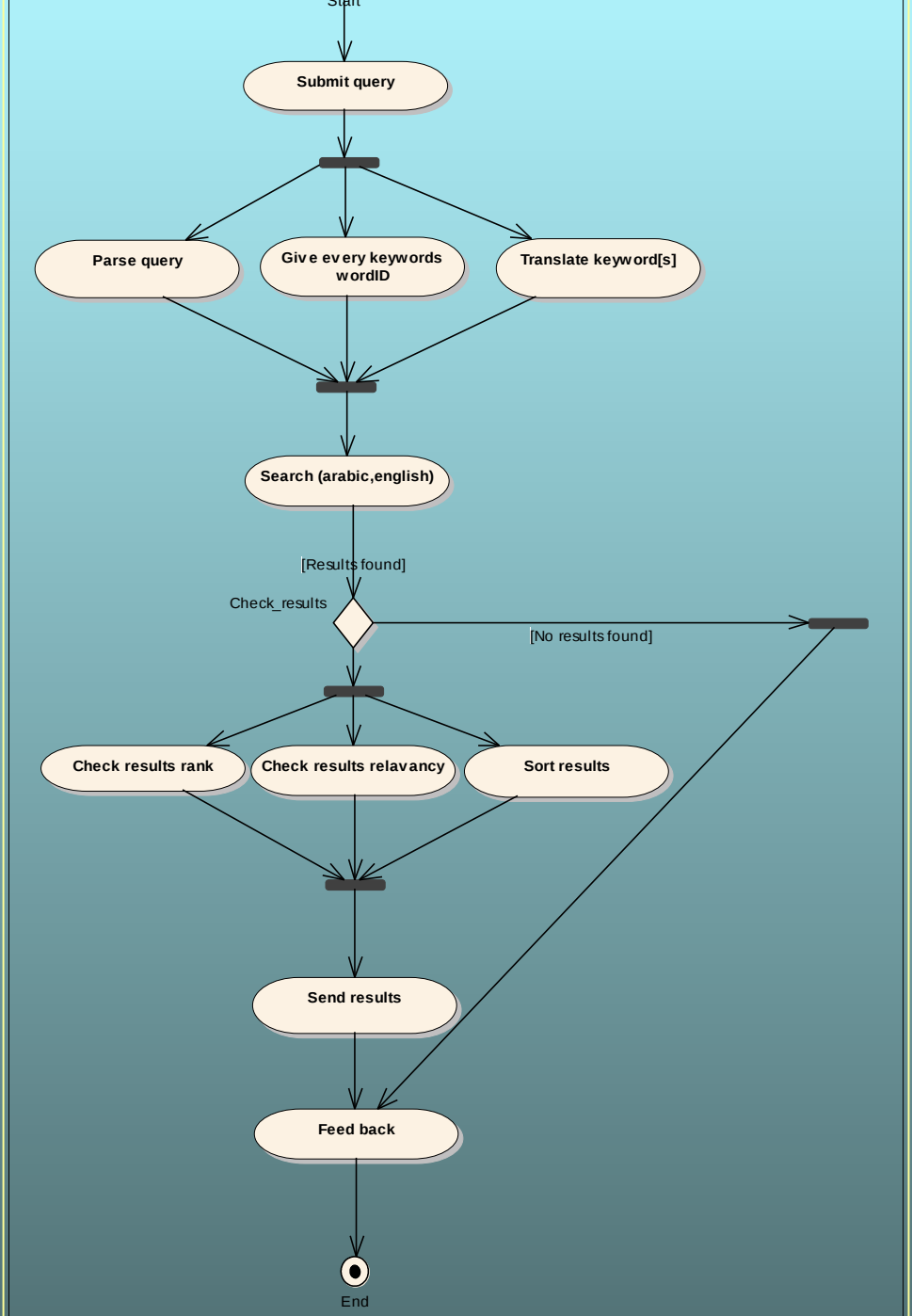


Downloads a Page Activity



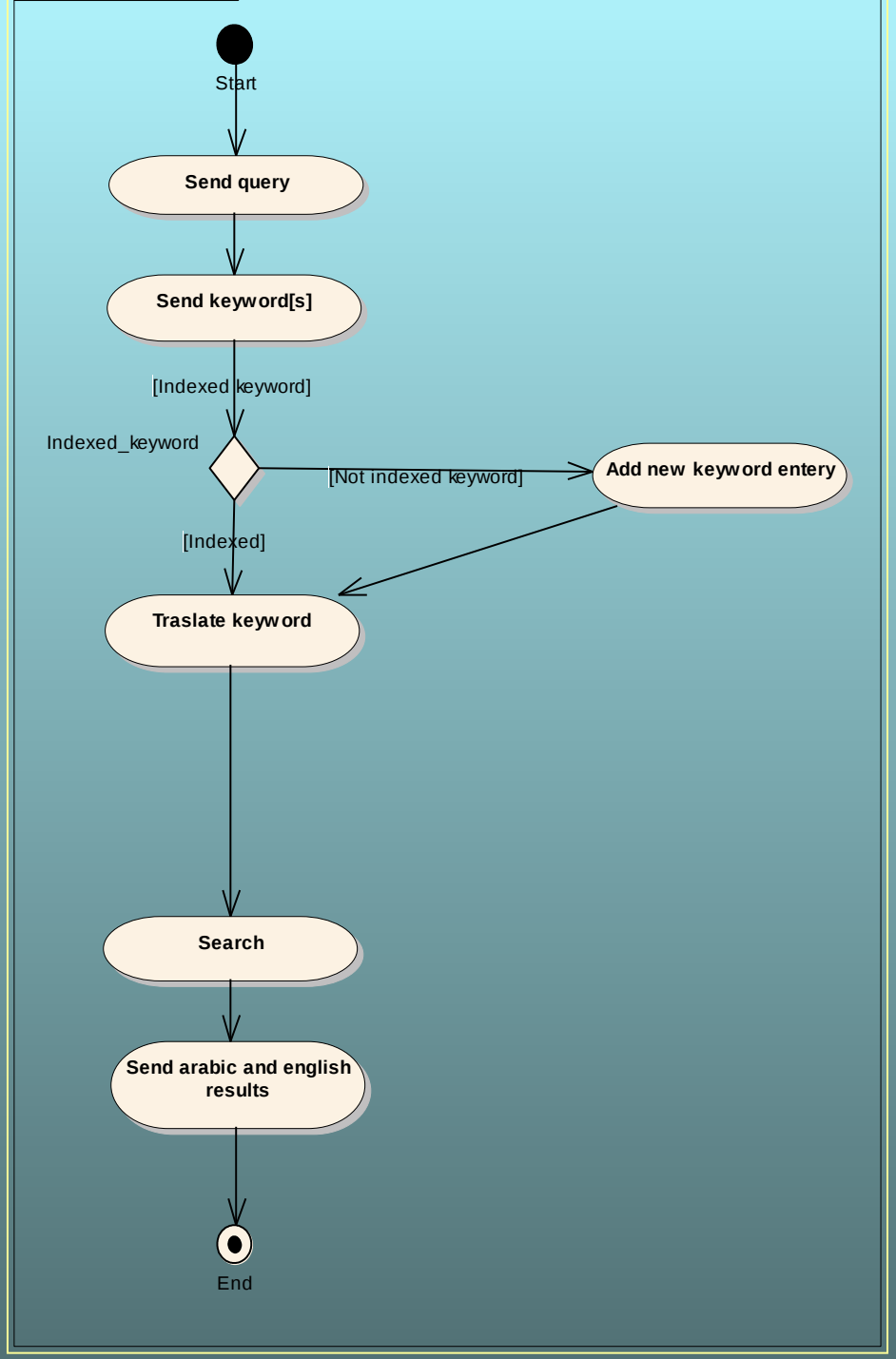
Indexes a Page Activity

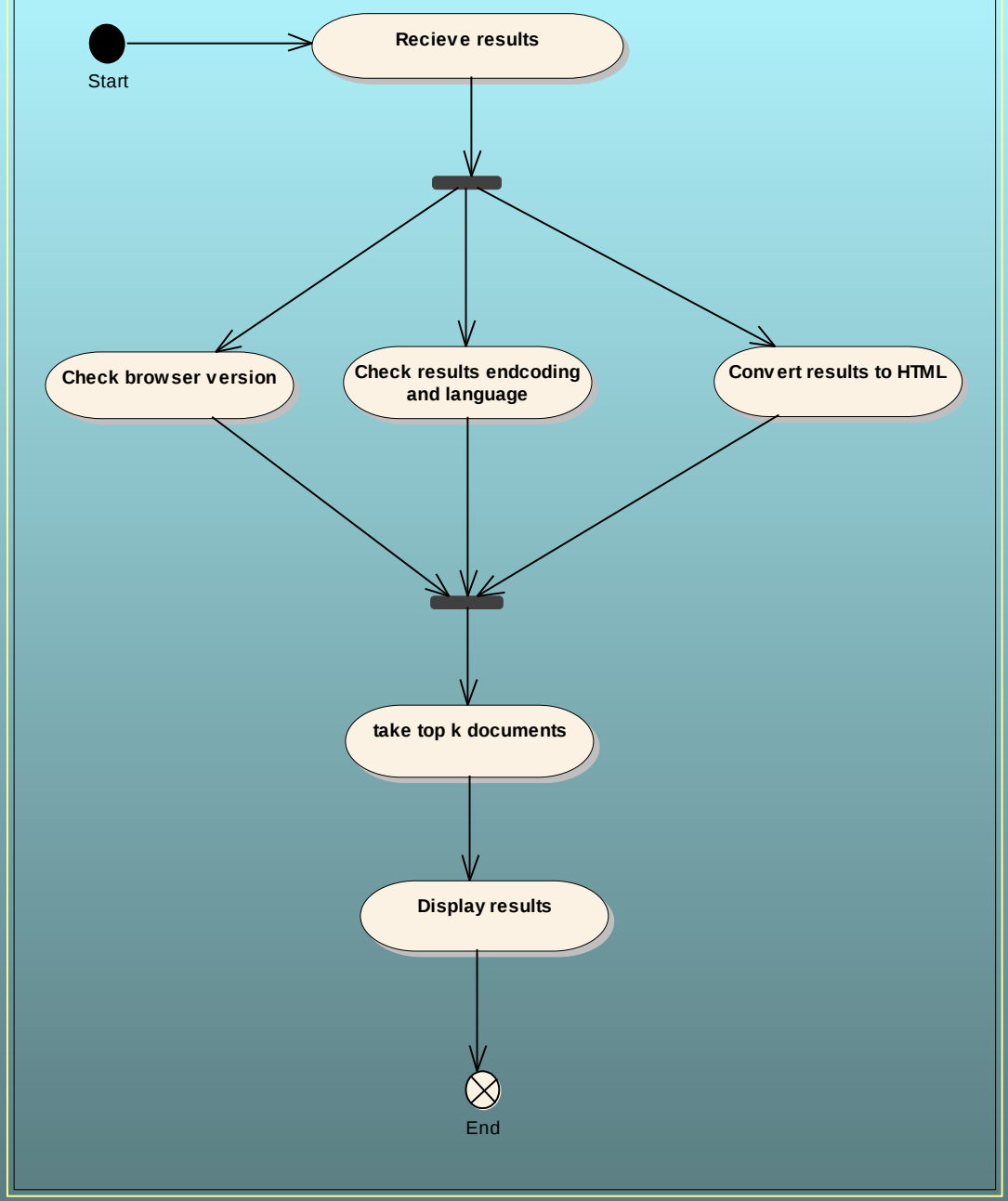




Processes Query Activity

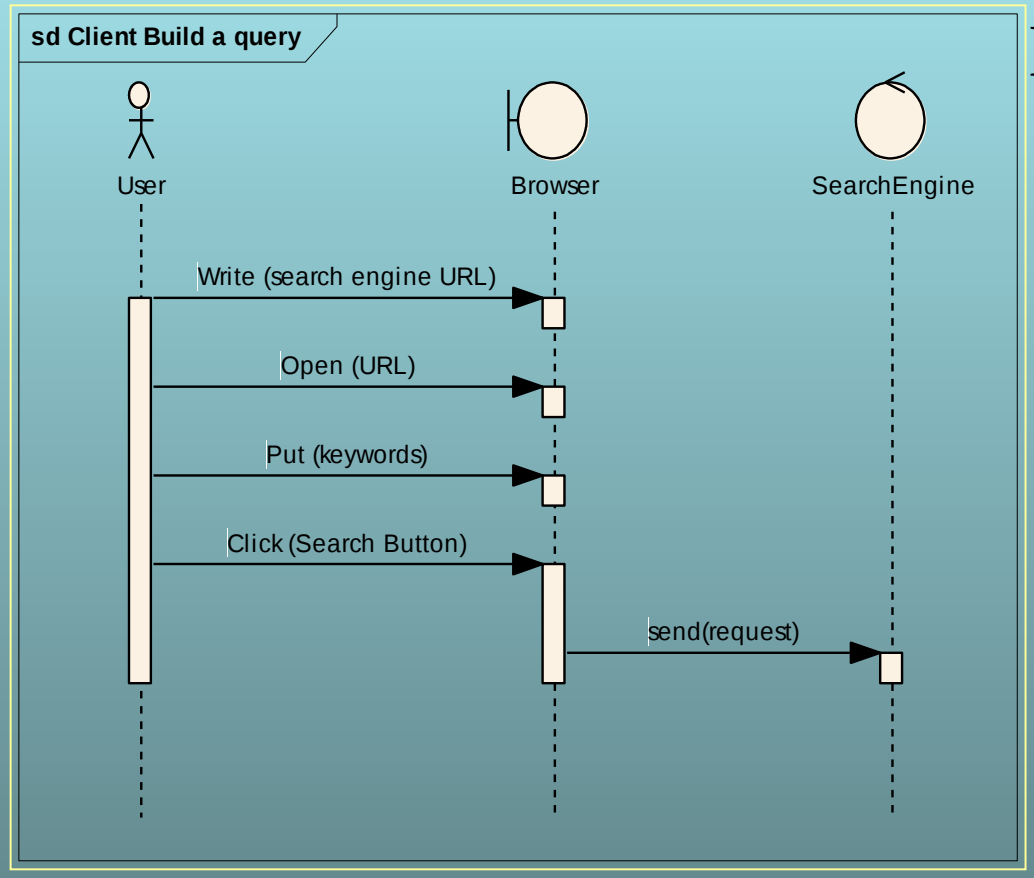
Translate Keyword[s] Activity



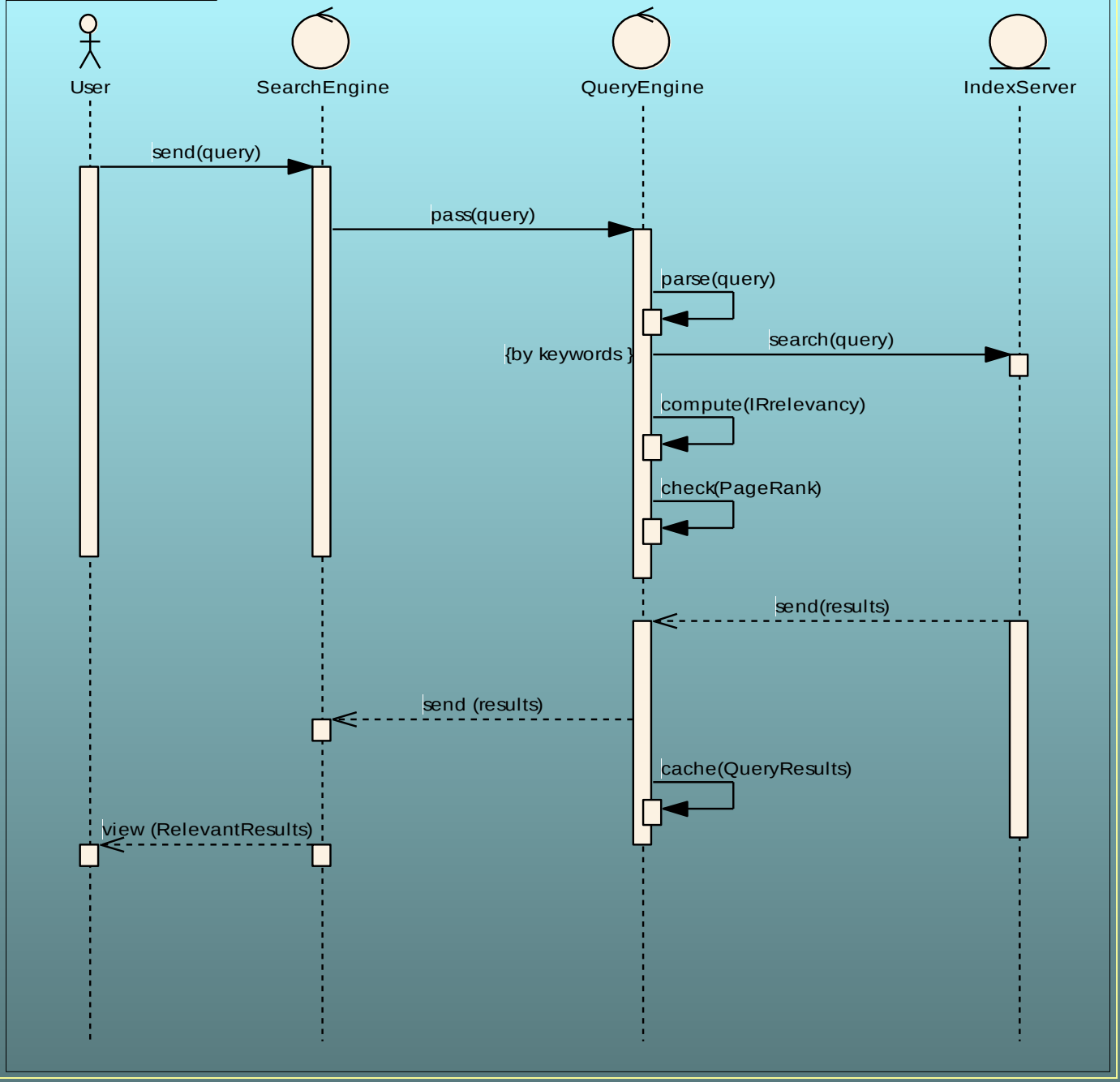


Display Results Activity

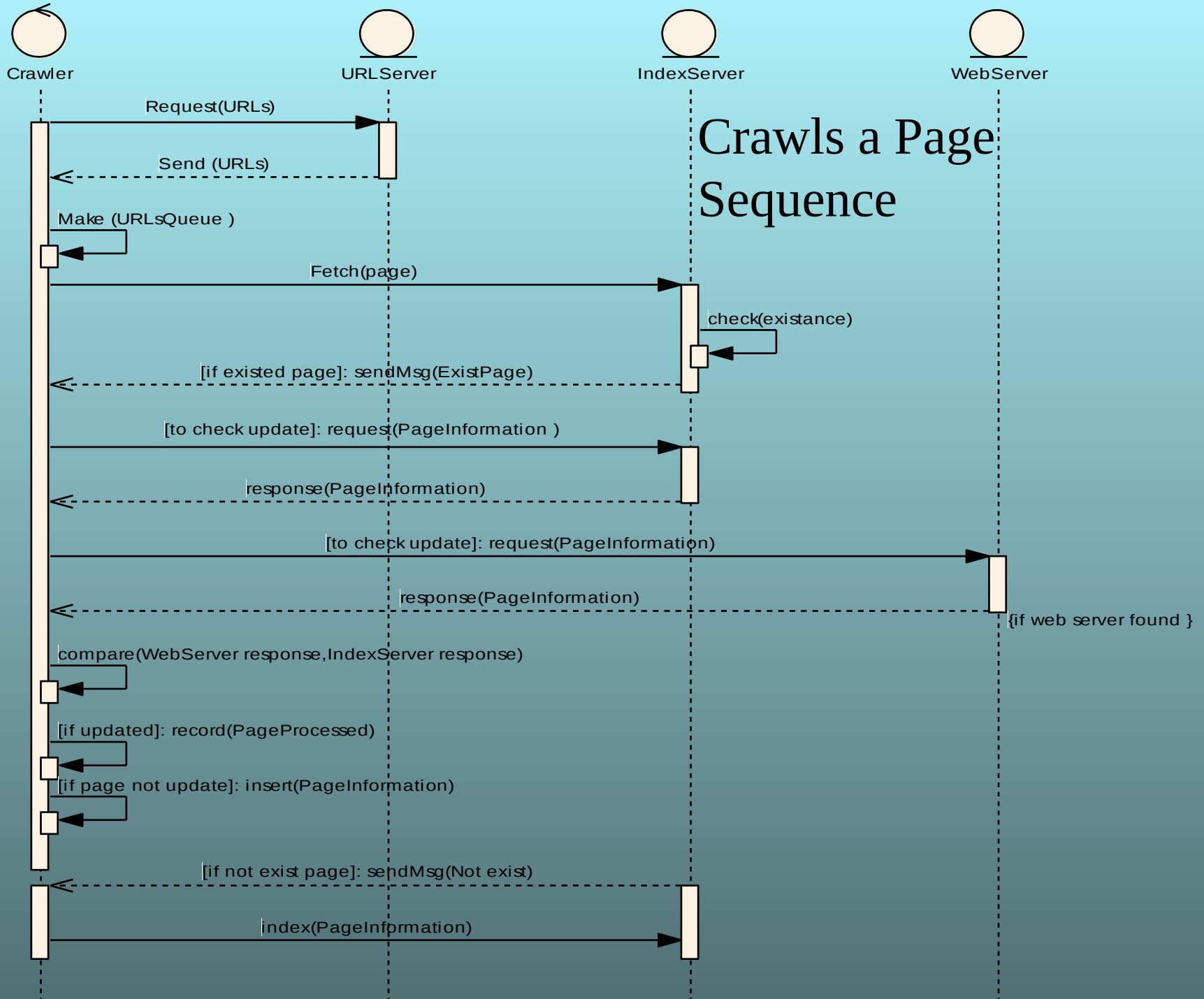
Build a Query Sequence

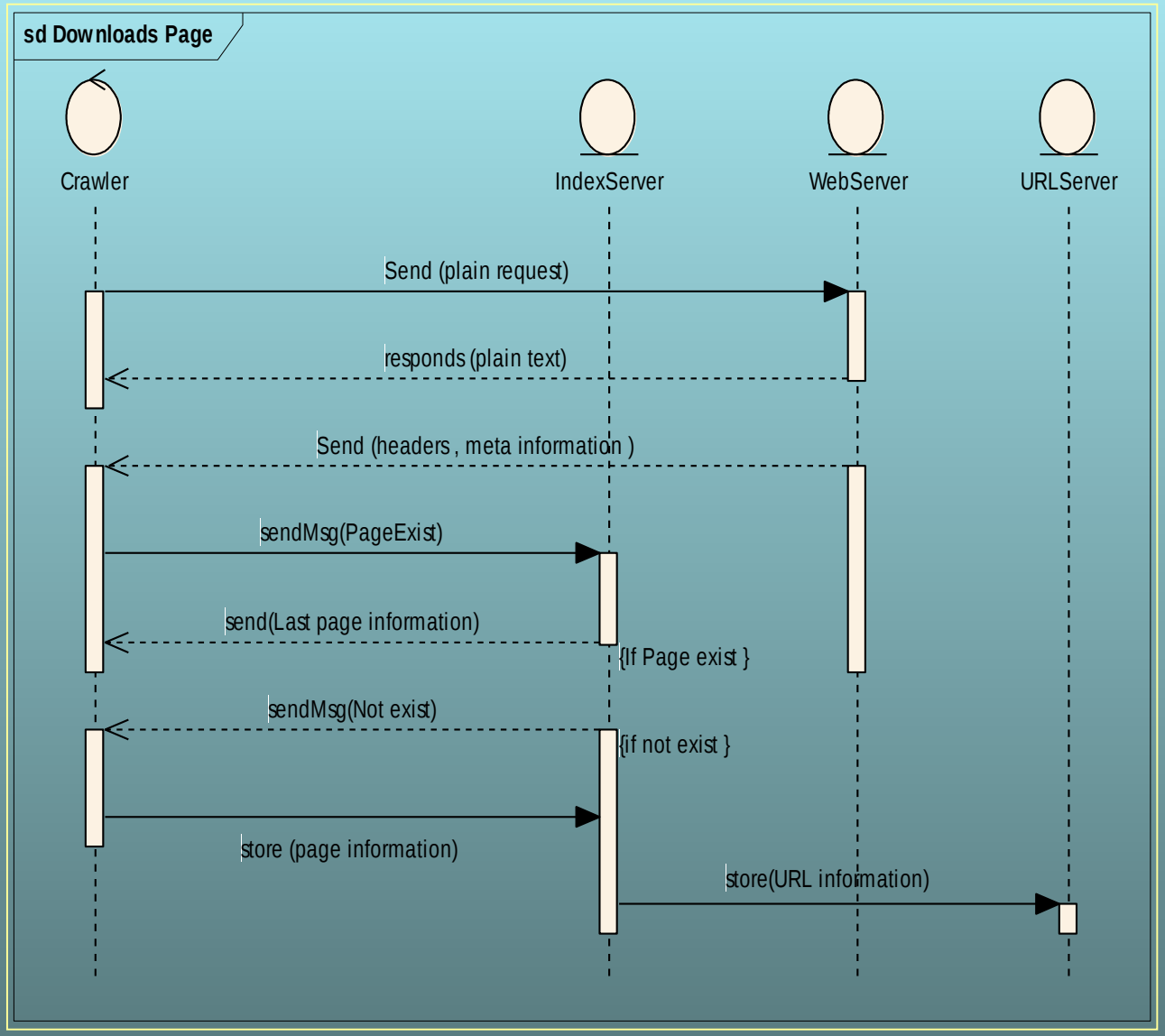


Check Relevancy Sequence



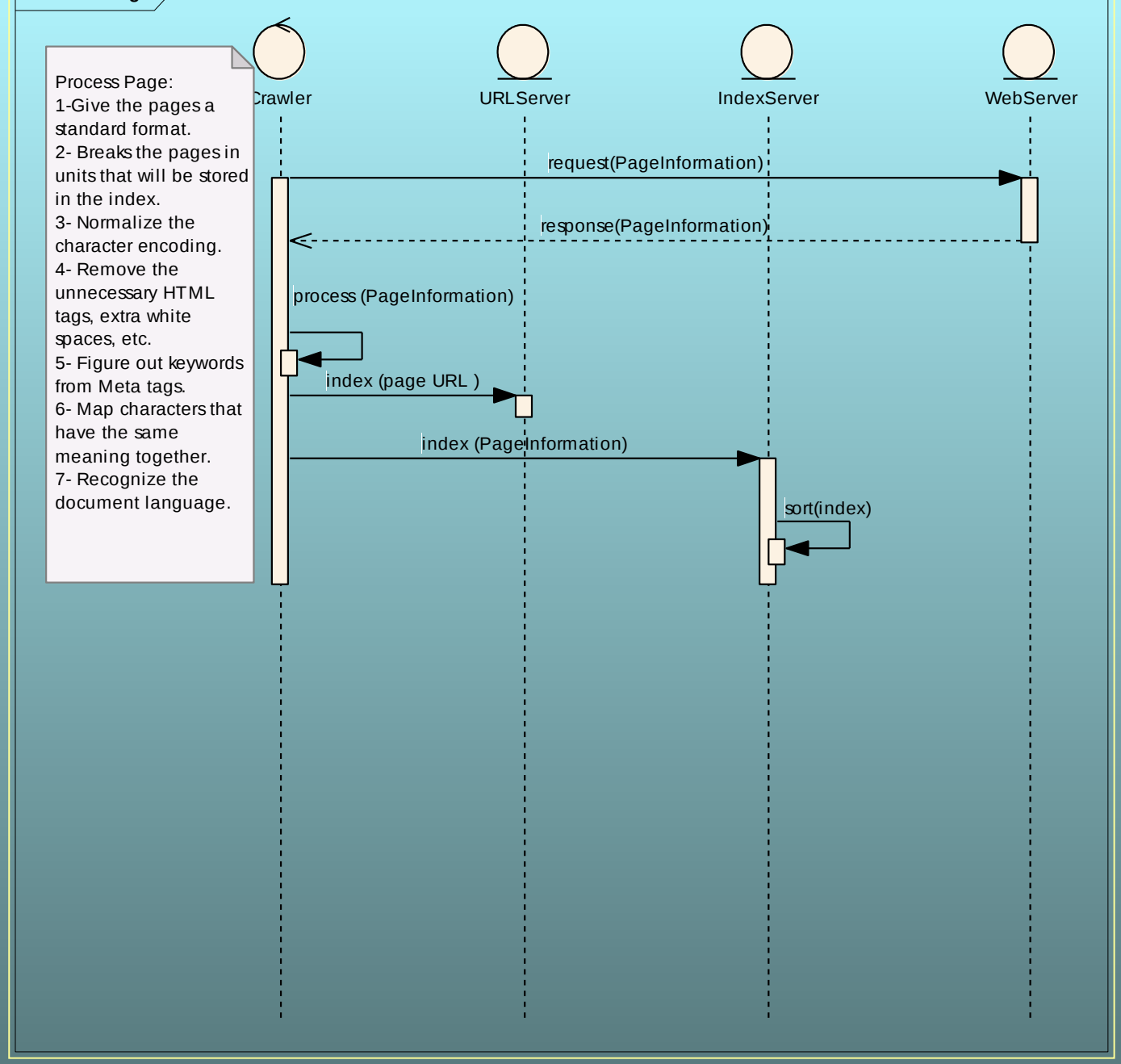
Crawls a Page Sequence





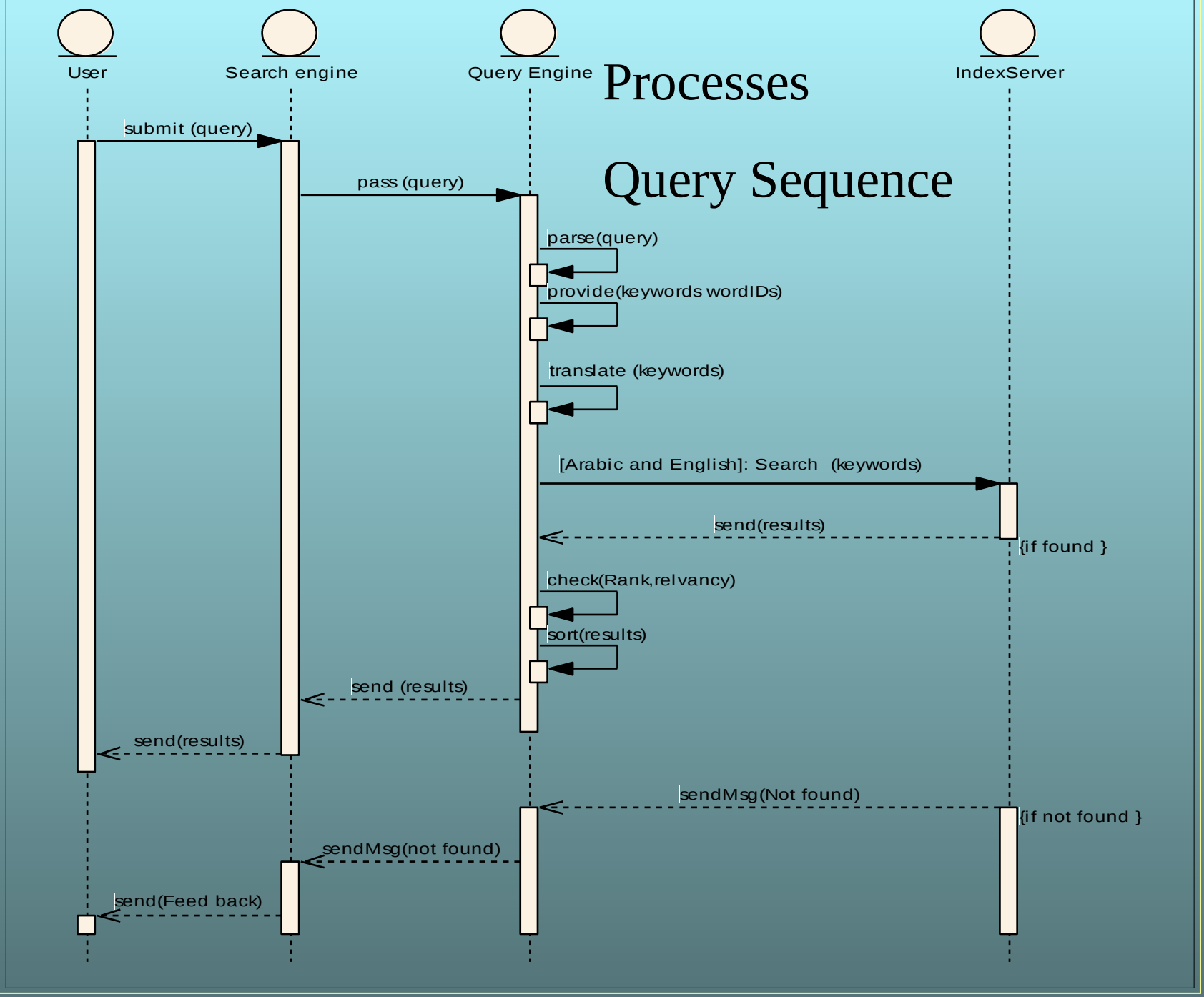
Downloads a Page Sequence

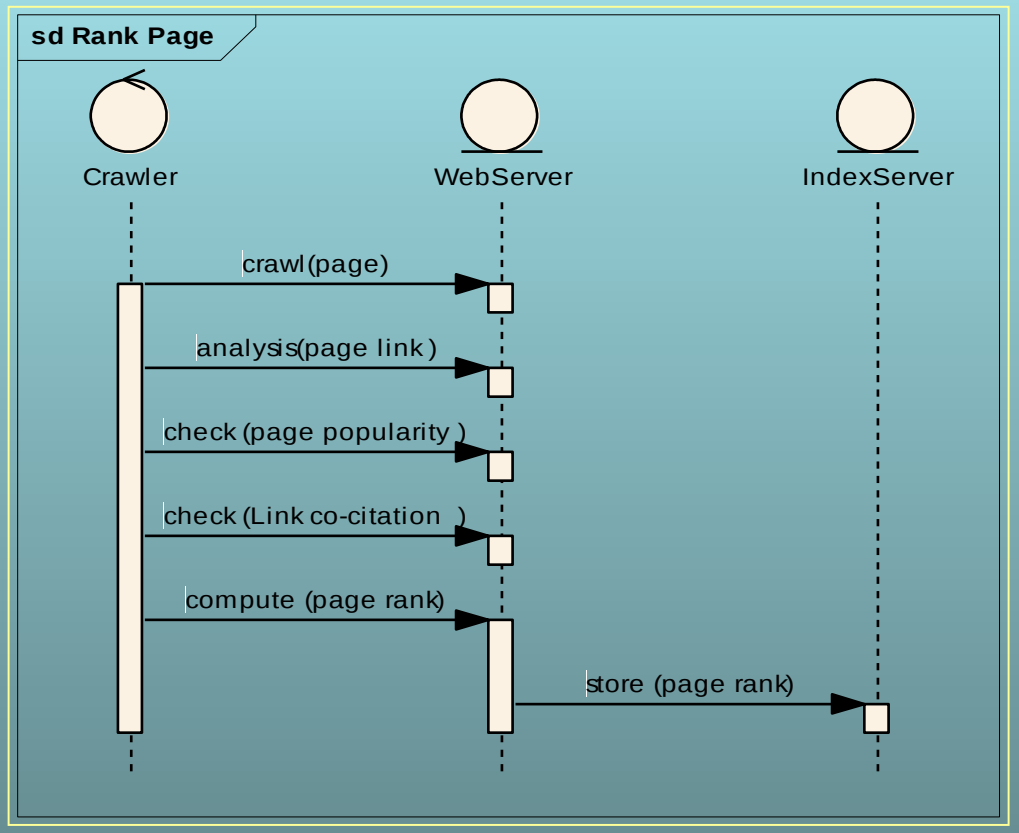
Indexes a Page Sequence



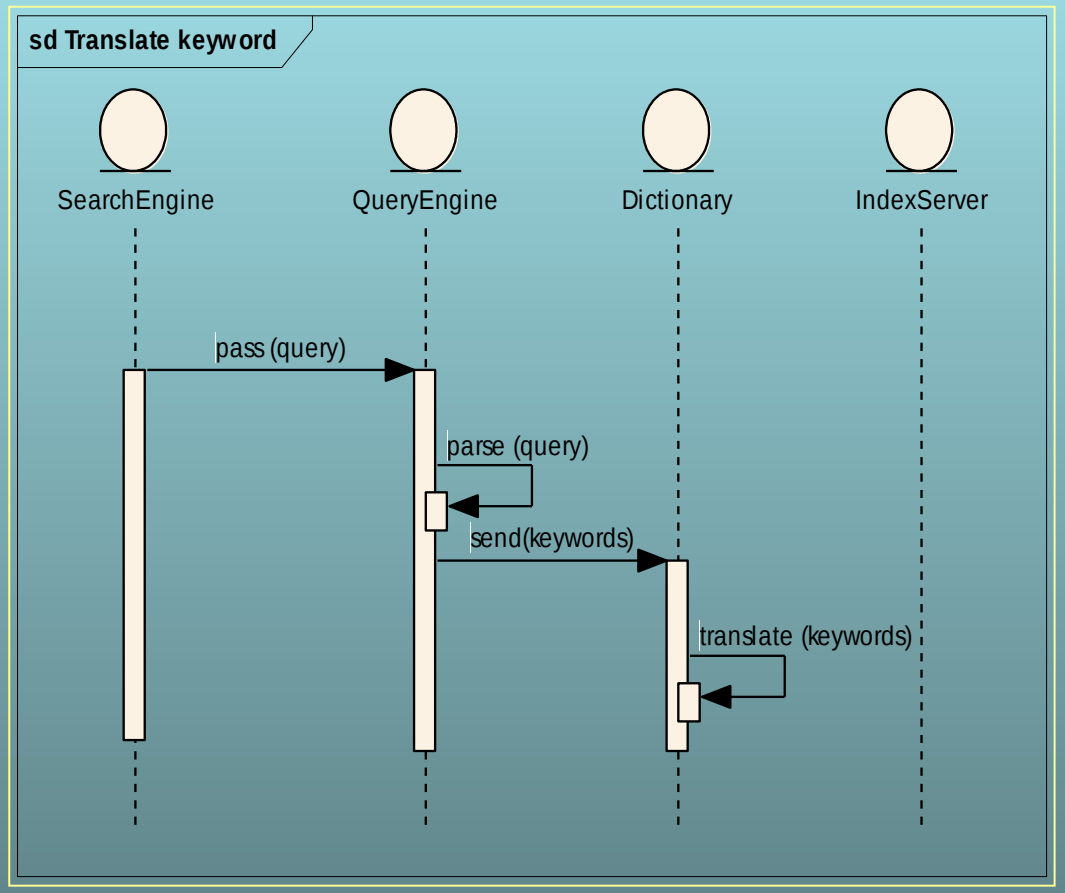
Processes

Query Sequence

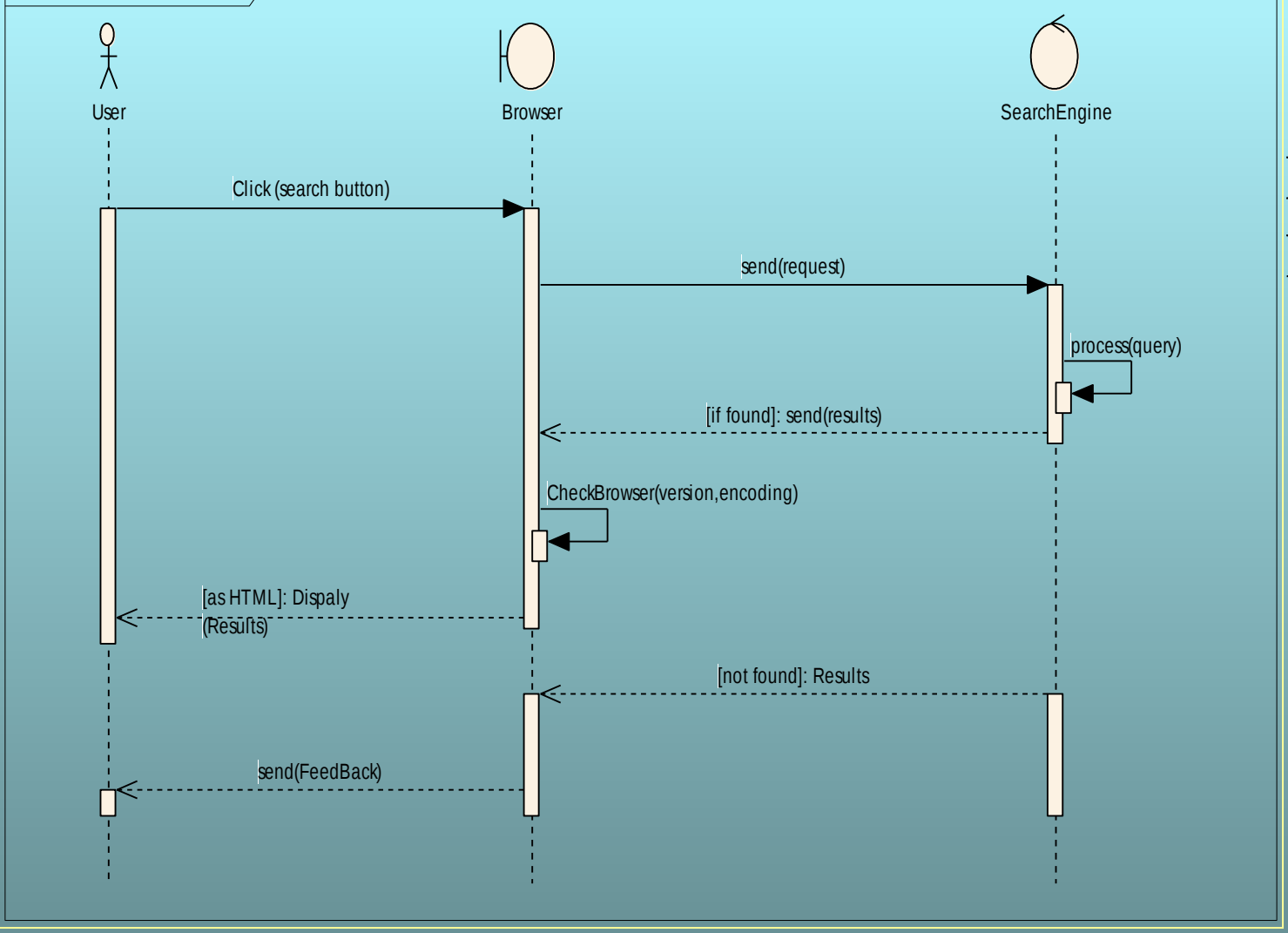




Ranks a Page Sequence



Translate Keyword Sequence



Display Results Sequence

Chapter 6: Case Study

Site description

As a case study of multilingual search, Sudan University of Science and Technology (SUST) journal site has taken to implement this search facilities.

The main site contains two separate Arabic and English sites to be searched separately, this achieved through these search keywords:

- Keyword[s].
- Article title.
- Author articles.
- Categories.

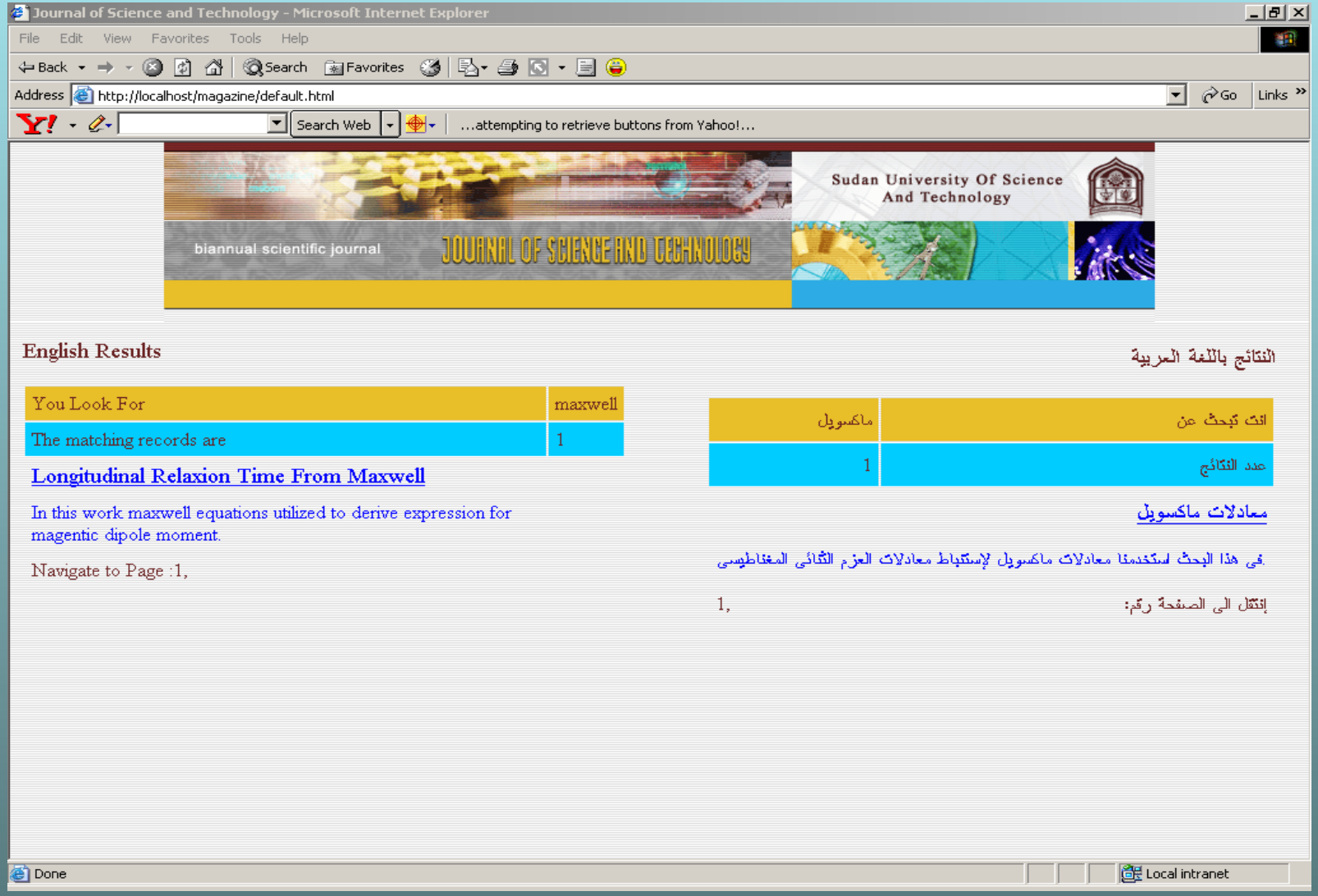
The New Site:

The user inputs his keyword (English keyword or Arabic keyword) in the search box. Then the query is processed in the steps below:

Connect to the database.

- Search through keywords table for user keywords.
- Find the corresponding meaning to the keyword.
- Send the Arabic keyword to Arabic table to search for it.
- Send the English keyword to English table to search for it
- Return the results as recordset.
- Display both Arabic results and English results.
- Divide them as five results per page.

Layout


Journal of Science and Technology - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://localhost/magazine/default.html
Go Links
Search Web
...attempting to retrieve buttons from Yahoo!...
Sudan University Of Science And Technology
biannual scientific journal
JOURNAL OF SCIENCE AND TECHNOLOGY
English Results
You Look For
maxwell
The matching records are
1
Longitudinal Relaxion Time From Maxwell
In this work maxwell equations utilized to derive expression for magentic dipole moment.
Navigate to Page :1,
النتائج باللغة العربية
انت تبحث عن
ماكسويل
عدد النتائج
1
معادلات ماكسويل
في هذا البحث استخدمنا معادلات ماكسويل لإستنباط معادلات العزم الثنائي المغناطيسي
إنتقل الى الصفحة رقم:
1,
Done
Local intranet

Chapter 7: Conclusion and Recommendations

- Although internet users differ in their nationalities and languages, search sites do not offer a multilingual search and most of them support one language search (English).
- The research is concerned with multilingual search process to enable users find information they need in both English and Arabic.
- The important points found on search engines area are crawling to the internet sites, indexes the resources found, give every resource a rank, and check information retrieval relevancy.
- In this research we have surveyed the components of search engines and have given an example of a multilingual search. We have implemented the search engine using Active Server Pages, SQL Server database, and ODBC.

Recommendations for Future Researches

Search engines are a non-limited field to study. Here are recommended areas to the future researches and studies:

- The process of query execution and optimization.
- Results caching to improve the search quality and response time.
- More through investigation of crawling algorithms and problems of invisible sites, dead links, and updated links.
- Ranking algorithms and information retrieval relevancy.
- Browsers encoding and versions and their influence on results displayed to users (frame pages, javascript, and dynamic pages, graphics).
- The implementation of a bilingual Arabic/English search engines.

رب اشرح لي صدري

والحمد لله رب العالمين