

Artificial Intelligence Information Retrieval(P76I900) -

Homework #1 Plus - P77131061 – 黃偉凱 (Huang, Wei-Kai)

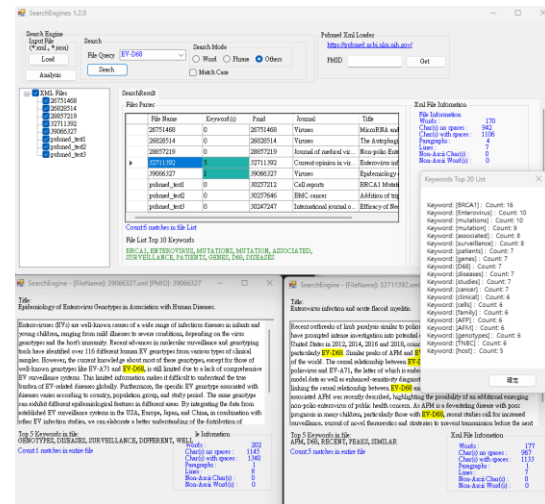
My Search Engine Application features

Application use GitHub manage

<https://github.com/rainsman-wk/ncku-113-P76I900>

This Homework Application Version : **V1.2** for demo in class

- Project Create by C# Windows Form (.Net Framework v4.8)
- Use MVP (**Model-View-Presenter**) Architecture design
- Simple View Explanation: **[MainForm]** / **[TopView]** / **[FileListView]** / **[ResultView]**
- All views as connect to **[ToolModel]** to do data exchange and trigger via **Event**.
- **File Paring** :Load feature in (.xml) with **xml element** search in <Abstract> and <AbstractText>
- **File Analyze** to List<string> of [Content], [Word] and [Sentence]
 - **The number of features use normal C# method to get counts (Incl. and excl. space)**
 - Non-Ascii count get by "[^\\u0000-\\u007F]" and "\\bW*" and "\\w*\\b"
- Each Load file store input data for view text search and save in class variants first.
- The Search feature including different search mode , the idea come from tool “Notepad++”
 - Search can provide in **Full Text Search** and **Selected Text Search** for each files
 - Search mode including Word(split by space) and search on **[Word List]**
 - Phrase search as same idea but search on **[Sentence List]**
 - Others search as search on **Regular expression** by **Regular expression**
- **The result have two block File Result** and **Content Result(Full Text)**
 - Files result can display all file analyze value and search result in remind file
 - Content Result can display all keywords and highlighted and give result counts



What is new in 2nd Homework #1 (Plus)

- More pop-up window feature for xml
- Add File information from Table to Text List (also for pop-up windows)
- Milit-Search in new windows and highlight
- New feature to search file keywords (Word Frequency)
- Use StopWord to get correct keywords for loading files
- Top 20 Word frequency information for loading file
- Change all search feature by regex

Plan to do (or extend feature for Homework#2)

- Save to Database (No which format have to use)
- Study method for Stemming
- ➔ Could use PorterStemmer to fix keyword issue (eq. mutations/mutation)