Tutorial 3 Exercise

Exercise 1

|  |  |
| --- | --- |
|  | Produce the Inverted Index using the template below. Use the Boolean Retrieval Model to perform the query of “Searching CCNA and SQLDev“, and rank the result based on relevance (word count). |

[File-3]

C

MYSQL

DB2

CCNA

IOS

SQLDev

CCNA

[File-5]

C

SQLDev

MYSQL

Oracle

Oracle

Java

SQLDev

DB2

DB2

MYSQL

MYSQL

[File-2]

Oracle

DB2

SQLDev

[File-4]

DB2

Java

Java

Oracle

CCNA

Php

Php

Php

SQLDev

MYSQL

[File-1]

SQLDev

CCNA

SQLDev

CCNA

CCNA

Template to be used

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Entry** | **Vocab** | **Posting** | | | |  |
| 1 | SQLDev | File-1(2) | File-5(2) |  |  |  |
| 2 | CCNA | File-1(1) | File-3(1) | File-4(1) |  |  |
| 3 | Oracle | File-2(1) | File-4(1) | File-5(2) |  |  |
| 4 | DB2 | File-2(1) | File-3(1) | File-4(1) |  |  |
| 5 | MYSQL | File-3(1) | File-5(1) |  |  |  |
| 6 | C | File-3(1) | File-5(1) |  |  |  |
| 7 | Java | File-4(4) | File-5(1) |  |  |  |
| 8 | PHP | File-4(1) |  |  |  |  |

**Searching CCNA and SQLDev**

1. Intersect 1st and 2nd row, and the result is [File-1] only.
2. Sum the frequency, we get [File-1](3)

Get the result of [File-1] with (3).

Exercise 2 - Case Study

iMedi is a private medical / healthcare group which provides out-patient medical clinic services to people in Hong Kong. The iMedi has 36 branches located in the 18 districts of Hong Kong and is serving around 30 thousand patients. Due to the ever-increasing workload and huge amount of patient data, the iMedi is considering building a centralized database for the 36 branches for their daily operations. You are the database consultant employed by the iMedi, and the iMedi’s senior management requests you to provide information for the questions below:

|  |  |
| --- | --- |
| **(a)** | **Explain what data and information are. Give two examples for data and information.**  Aws:  Data is a set of discrete and objective facts about events. It is usefully described as structured records of transactions.  Example: personal (e.g. HKID#, name, age, address, bank account numbers, etc.) and job/salary (e.g. rank, mater-pay-scale points, etc.)  Information is structured data to make sense and organized to some purposes.  Example: Data would be transformed into information when its creator adds meaning. |
| **(b)** | **Explain why data and information are important corporate resources. From the above case study.**  Aws:  Data and Information are important corporate resources because |
| **(c)** | **Based on the iMedi, please identify the three main advantages and three disadvantages of using database.**  Aws:  Advantage:  1) Centralize all data  2) Structure all data  3) Have Recoverability  Disadvantage:  1) Initial training required for all programmer and users  2) Hardware and software can be very costly  3) Maintenance cost can be height |

|  |  |
| --- | --- |
| **(d)** | **Based on the iMedi, please identify the three main advantages and three disadvantages of using Information Retrieval (IR).**  Aws:  Advantage:  1) Can search in free word (can use by someone not professional in computer).  2) by relevance or by meaning but no need exact match (Fuzzy search)  3) Can be scattered in millions of computers (Distributed computing, data protection)  Disadvantage:  1) Imprecise and might change from time by time  2) unstructured data may become messy  3) large Scale (not suitable for small company) |
| **(e)** | **Based on the DIKW framework, discuss how information can be generated from data, knowledge from information, etc.**  Aws:  Information is organized data represented to generate relations.  Data become information after contextualized, categorized, calculated, corrected and condensed.  Knowledge is contextual, relevant, actionable information from observed patterns of information.  Information become knowledge after comparison, consequence, connections and conversation. |