|  |
| --- |
|  |
| HW3 - Indexing |
|  |
| PB15111604金泽文 |

[**Web Information Processing and Applications**](http://staff.ustc.edu.cn/~jpq/courses/webinfo.html)

HW3 - Indexing

PB15111604金泽文

1. 考虑下面的文档：

**Doc 1**    prediction of whole country sales   
    **Doc 2**    country sales rise in July   
    **Doc 3**    decrease in home sales in June   
    **Doc 4**    July country sales rise prediction

* 1. term-document关联矩阵

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 |
| Country | **1** | **1** | **0** | **1** |
| Decrease | **0** | **0** | **1** | **0** |
| Home | **0** | **0** | **1** | **0** |
| In | **0** | **1** | **2** | **0** |
| July | **0** | **1** | **0** | **1** |
| June | **0** | **0** | **1** | **0** |
| Of | **1** | **0** | **0** | **0** |
| prediction | **1** | **0** | **0** | **1** |
| rise | **0** | **1** | **0** | **1** |
| sales | **1** | **1** | **1** | **1** |
| whole | **1** | **0** | **0** | **0** |

* 1. 倒排索引

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | doc.freq. |  | postings list | | | |
| country | **3** | **=>** | **1 [freq.=1]** | **2 [freq.=1]** | **4 [freq.=1]** |  |
| decrease | **1** | **=>** | **3 [freq.=1]** |  |  |  |
| Home | **1** | **=>** | **3 [freq.=1]** |  |  |  |
| In | **2** | **=>** | **2 [freq.=1]** | **3 [freq.=2]** |  |  |
| July | **2** | **=>** | **2 [freq.=1]** | **4 [freq.=1]** |  |  |
| June | **3** | **=>** | **3 [freq.=1]** |  |  |  |
| Of | **1** | **=>** | **1 [freq.=1]** |  |  |  |
| prediction | **2** | **=>** | **1 [freq.=1]** | **4 [freq.=1]** |  |  |
| rise | **2** | **=>** | **2 [freq.=1]** | **4 [freq.=1]** |  |  |
| sales | **4** | **=>** | **1 [freq.=1]** | **2 [freq.=1]** | **3 [freq.=1]** | **4 [freq.=1]** |
| whole | **1** | **=>** | **1 [freq.=1]** |  |  |  |