**Wikipedia Matrix Analyses**

**Population**

Pour répondre aux question ont été analysés 1426 *PCMs*

**Nom de *features* les plus fréquentes**

|  |  |  |
| --- | --- | --- |
| **Value** | **Fréquence absolue** | **Fréquence relative (%)** |
|  | 420 | 2,4494 |
| yes | 385 | 2,2453 |
| model | 179 | 1,0439 |
| name | 166 | 0,9681 |
| license | 149 | 0,8690 |
| ? | 143 | 0,8340 |
| notes | 133 | 0,7756 |
| code name | 131 | 0,7640 |
| bus width (bit) | 115 | 0,6707 |
| bandwidth (gb/s) | 112 | 0,6532 |
| no | 112 | 0,6532 |
| bus interface | 111 | 0,6473 |
| fab (nm) | 106 | 0,6182 |
| memory (mb) | 106 | 0,6182 |

**Table 1. Les 15 *features* le plus frequentes**

**Nom de *produits* les plus fréquentes**

|  |  |  |
| --- | --- | --- |
| **Value** | **Fréquence absolue** | **Fréquence relative (%)** |
| p3 | 73 | 0,26597 |
| p1 | 72 | 0,26232 |
| p2 | 72 | 0,26232 |
| p4 | 70 | 0,25504 |
| p6 | 70 | 0,25504 |
| p5 | 69 | 0,25139 |
|  | 68 | 0,24775 |
| p7 | 68 | 0,24775 |
| p8 | 66 | 0,24046 |
| p9 | 66 | 0,24046 |
| p10 | 65 | 0,23682 |
| p11 | 61 | 0,22225 |
| p12 | 58 | 0,21132 |
| p13 | 58 | 0,21132 |
| p14 | 57 | 0,20767 |

**Table 2. Les 15 *produits* le plus frequentes**

**Taille de matrices**

|  |  |
| --- | --- |
| **Quantité de PCMs** | 1426 |
| **Taille moyenne de une PCM** | 205,98 |
| **Taille moyenne des features dans en PCM** | 11,87 |
| **Taille moyenne des produits dans en PCM** | 19,38 |

**Table 3. Information generale sur la taille de matrices**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nom** | **Taille *features*** | **Taille *produits*** | **Taille total** |
| Comparison\_of\_Chinese\_romanization\_systems\_0.pcm | 11 | 411 | 4521 |
| Comparison\_of\_layout\_engines\_(Document\_Object\_Model)\_3.pcm | 9 | 355 | 3195 |
| Comparison\_of\_e-book\_readers\_3.pcm | 35 | 75 | 2625 |
| Comparison\_of\_orbital\_launch\_systems\_0.pcm | 10 | 238 | 2380 |
| Comparison\_of\_layout\_engines\_(Cascading\_Style\_Sheets)\_2.pcm | 392 | 6 | 2352 |
| Comparison\_of\_graphics\_file\_formats\_0.pcm | 9 | 235 | 2115 |
| Comparison\_of\_machine\_translation\_applications\_1.pcm | 16 | 124 | 1984 |
| Comparison\_of\_antivirus\_software\_0.pcm | 26 | 76 | 1976 |
| Comparison\_of\_AMD\_processors\_0.pcm | 15 | 128 | 1920 |
| Comparison\_of\_netbooks\_0.pcm | 22 | 85 | 1870 |
| Comparison\_of\_Dewey\_and\_Library\_of\_Congress\_subject\_classification\_0.pcm | 4 | 456 | 1824 |
| Comparison\_of\_online\_backup\_services\_0.pcm | 30 | 60 | 1800 |
| Comparison\_of\_Exchange\_ActiveSync\_clients\_0.pcm | 99 | 18 | 1782 |
| Comparison\_of\_IOC,\_FIFA,\_and\_ISO\_3166\_country\_codes\_0.pcm | 7 | 254 | 1778 |
| Comparison\_of\_XMPP\_server\_software\_1.pcm | 13 | 129 | 1677 |

**Table 4. Les 15 matrices le plus grandes**

**Valeurs de *cells* les plus fréquentes**

|  |  |  |
| --- | --- | --- |
| **Value** | **Fréquence absolue** | **Fréquence relative (%)** |
| yes | 50810 | 17,3962 |
| no | 42815 | 14,6589 |
|  | 26373 | 9,0295 |
| ? | 16674 | 5,7088 |
| unknown | 5180 | 1,7735 |
| 1.0 | 1972 | 0,6752 |
| proprietary | 1901 | 0,6509 |
| n/a | 1469 | 0,5030 |
| gpl | 699 | 0,2393 |
| free | 687 | 0,2352 |
| partial | 660 | 0,2260 |
| 128 | 566 | 0,1938 |
| 2.0 | 511 | 0,1750 |
| 1 | 461 | 0,1578 |
| 256 | 433 | 0,1482 |

**Table 5. Les 15 *cells* le plus frequentes**

**Type de valeurs les plus fréquentes**

|  |  |  |
| --- | --- | --- |
| **Type de donnée** | **Fréquence absolue** | **Fréquence relative (%)** |
| booleanvalueimpl | 93764 | 32,52 |
| stringvalueimpl | 86972 | 30,16 |
| partialimpl | 660 | 0,23 |
| realvalueimpl | 11126 | 3,86 |
| multipleimpl | 18466 | 6,40 |
| notavailableimpl | 50487 | 17,51 |
| integervalueimpl | 18926 | 6,56 |
| conditionalimpl | 7956 | 2,76 |

**Table 5. *Cells* partype**

**Valeurs de cellule qui sont le plus souvent ensemble**

|  |  |  |
| --- | --- | --- |
| **Pairs** | **Fréquence absolue** | **Fréquence relative (%)** |
| **[yes, no]** | **3554** | **4%** |
| **[yes, ?]** | **1310** | **1,6%** |
| **[No, ?]** | **805** | **0,09%** |
| **[yes, null]** | **255** | **0,03%** |
| **[ ,yes]** | **231** | **0,02%** |
| **[unknown, yes]** | **217** | **0,02%** |
| **[ ,No]** | **210** | **0,002%** |
| **[unknown, no]** | **197** | **0,002%** |
| **[yes, partial]** | **196** | **0,002%** |
| **[no , nul l]** | **179** | **0,002%** |

**Correlation type-feature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Features | Type | Quantite | Type | Quantité |
| Rank | IntegerValueImpl | 181 | NotAvailableImpl | 26 |
| Portuguese | MultipleImpl | 3 | StringValueImpl | 13 |
| English | MultipleImpl | 4 | StringValueImpl | 12 |

**Table 6. Features et type de cell**

On constate que le type prédominant de cell détermine la nature des features. Quantitatif, Qualitatif…

**Homogénéité de colonnes**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Features | Type | Quantite | Type | Quantité |
| Rank | IntegerValueImpl | 181 | NotAvailableImpl | 26 |
| Portuguese | MultipleImpl | 3 | StringValueImpl | 13 |
| English | MultipleImpl | 4 | StringValueImpl | 12 |

**Table 7. Features et type de cell**

On constate que des colonnes on plusieurs type de diffèrent donc il n’y a pas d’homogénéité.

**Similarité de matrices**

1,Comparison\_between\_Argentine\_provinces\_and\_countries\_by\_GDP\_(PPP)\_per\_capita , Comparison\_between\_Esperanto\_and\_Ido - Morphology , 3 , 6 , 207 , 12 , 0 , 0 , different  
2,Comparison\_between\_Argentine\_provinces\_and\_countries\_by\_GDP\_(PPP)\_per\_capita , Comparison\_between\_Esperanto\_and\_Ido - Vocabulary , 3 , 14 , 207 , 16 , 0 , 0 , different

La comparaison de PCM nous donc les information suivante :

|  |  |
| --- | --- |
| Titre | Type |
| Nom premier PCM (pcm1) | String |
| Nom deuxième PCM (pcm2) | String |
| Nombre de freatures premier PCM | Numérique |
| Nombre de freatures deuxième PCM | Numérique |
| Nombre de produits premier PCM | Numérique |
| Nombre de produits deuxième PCM | Numérique |
| Nombre de Features en commun | Numérique |
| Nombre de produits en commun sur les même features | Numérique |
| Résultat de la comparaison | Similaire ; Diffèrent ; pcm1 est inclut dans pcm2 ou pcm2 est inclut pcm1 |

**Table 7. Structure de Comparaison**

**Procédure d’extraction défectueuse**

Dans le cas suivante nous avons trouvé procédure d’extraction défectueuse

* *Features* qui ne contient pas de cellules
* Sautes de ligne dedans les *cells*
* Il n’a pas information sur les images trouvent dedans les cells