**Penjelasan Atau Overview Beberapa Minor Task Dalam Web Content Mining (WCM)**

1. **Pembahasan Task**
2. **Information Visualization**

Information visualization menggunakan feature extraction dan key indexing untuk membuat grafik atau bentuk visualisasi lain yang dapat merepresentasikan data. Dengan proses visualisasi, dokumen yang memiliki kesamaan dapat ditemukan dengan lebih cepat. Kumpulan text yang tidak dapat ditampilkan secara langsung dengan gambar dapat di buat menjadi visual hierarchy atau map yang dapat ditelusuri. Penggambaran ini akan memudahkan user untuk menganalisa konten dari dokumen. Gambar graph atau map yang dibuat dapat diperbesar maupun dipecah menjadi sub-map lebih kecil. Teknik ini berguna untuk menemukan topic yang berkaitan dalam dokumen yang berjumlah sangat banyak atau berukuran sangat besar.

Sumber :

International Journal of Computer Applications 47 : Web Content Mining Techniques: A Survey

IOSR Journal of Computer Engineering : Web Content Mining: Tool, Technique & Concept

1. **Page Content Mining**

Page Content Mining merupakan salah satu cara structured data extraction yang menggunakan page ranking system pada hasil pencarian page. Ranking system bisa didapatkan dari rank algorithm milik google saat hasil search page di load. Proses Page Content Mining akan membandingkan ranking content, sehingga halaman website dapat diklasifikasikan.

Sumber :

International Journal of Computer Applications 47 : Web Content Mining Techniques: A Survey

IOSR Journal of Computer Engineering : Web Content Mining: Tool, Technique & Concept

1. **Object Exchange Model**

Informasi yang penting akan diekstrak menggunakan semi-structured data mining technique dan informasi penting tersebut akan disisipkan menjadi beberapa kelompok atau grup dan disimpan pada object exchange mode (OEM). OEM ini membantu user memahami struktur informasi dengan lebih akurat. Penggunaan OEM lebih sering dipakai untuk konten web yang memiliki banyak variasi atau heterogeneous.

Sumber :

International Journal of Computer Applications 47 : Web Content Mining Techniques: A Survey

IOSR Journal of Computer Engineering : Web Content Mining: Tool, Technique & Concept

1. **Wrapper Generation**

Wrapper generation memberi informasi berdasarkan sumber data diambil. Secara umum, halaman website akan di ranking secara langsung oleh search engine, kemudian data tersebut diterima sesuai query pada halaman website dan diurutkan berdasar ranking halaman tersebut. Query tersebut akan menentukan tipe output yang diberikan, di mana wrapper juga akan menambahkan informasi seperti domain, statistic dan index look up dari halaman tersebut.

Sumber :

International Journal of Computer Applications 47 : Web Content Mining Techniques: A Survey

IOSR Journal of Computer Engineering : Web Content Mining: Tool, Technique & Concept

*\*(Jujur pak, dari pemahaman saya yang mungkin masih terlalu dasar ini. Saya masih agak bingung beda wrapper generation dengan page content mining. Apa ranking di page content hanya mengukur content sedangkan wrapper ini berdasar query yang di wrap? Atau berdasar ranking pada search engine nya sendiri? Mungkin saya yang membacanya kurang, atau penyerapan informasi saya yang tidak bagus)*

1. **Web Data Extraction Language**

Web data extraction adalah sebuah proses mengkonversi resources dari halaman website menjadi sebuah format yang berstruktur. Misalkan sebuah halaman website mendeskripsikan informasi mengenai beberapa perusahaan, maka web data extraction akan mengubah data HTML pada halaman website menjadi struktur data yang dapat diproses oleh komputer. Seperti relational database table. Tujuan utama dari data extraction adalah membuat informasi data bisa diubah menjadi informasi yang lebih berguna seperti report analytic.

Sumber :

Encyclopedia of Database Systems : Languages for Web Data Extraction

IOSR Journal of Computer Engineering : Web Content Mining: Tool, Technique & Concept

1. **Daftar Pustaka**
2. Johnson, Faustina dan Gupta, Santosh Kumar. 2012. *Web Content Mining Techniques: A Survey*. International Journal of Computer Applications 47. DOI:10.5120/7236-0266
3. Sinha, Anil & Raj, Nidhi & Haque, Shameemul & Haque, Md dan N.K.Singh,. 2016. *Web Content Mining: Tool, Technique & Concept.* IOSR Journal of Computer Engineering. DOI:10.9790/0661-1806065760
4. Kushmerick, N. 2009. *Languages for Web Data Extraction*. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA. DOI:10.1007/978-0-387-39940-9\_1156