1. Motivation

A user is exploring Wikipedia articles using keyword query...

Facets!

Then using two facets intersection to find Wikipedia article:

Facets in Wikipedia

US action film

Facetedpedia: Enabling Query-Dependent Faceted Search for Wikipedia

http://idir.uta.edu/facetedpedia

Ning Yan, Chengkai Li, Senjuti B. Roy, Rakesh Ramegowda, Gautam Das

3. Facetedpedia Architecture

Preprocessing
- remove redirect pages and categories
- remove cycles from Wikipedia category hierarchy

Facet Discovery
- multi-thread background daemon loads preprocessed data into memory
- discover faceted interface per user session

Categorization
- assign Wikipedia articles to 80 domains

Facetedpedia Web GUI
- AJAX based Web GUI

4. Facet Discovery

The problem: discover the best k-facet interface

Step1: Retrieve attribute articles and build RCH
Step2: Rank single facets based on navigational cost
Step3: Select multiple facets using hill-climbing method

5. Facet Ranking

Single-Facet Ranking

\[ \text{cost}(F_i) = \frac{1}{7} \left( \sum \text{cost(path } L_{p\in(p_0, \ldots, p_7)} \right) + \text{penalty} \times \#L'_{p\in(p_1)} \]

Multi-Facet Selection

1. Sum of single facet cost is small
2. Overlapping of k-facet is small (Jaccard Similarity)

6. System Configuration

Data Set:
- Wikipedia data dump July 24th 2008
- 2.4M target articles, 110M attribute article hyperlinks
- 330K categories, 730K category links in category hierarchy

References
