

Facetedpedia: Dynamic Generation of Query-Dependent Faceted Interfaces for Wikipedia

Chengkai Li, **Ning Yan**, Senjuti B. Roy,
Lekhendo Lisham, Gautam Das

Department of Computer Science and Engineering
University of Texas at Arlington

June 18, 2010

WHAT IS A FACETED INTERFACE

ning yan [log out] 0 Items (\$ 0.00) MY ACCOUNT

MY NEWEGG
REVIEWS
HELP & INFO

SEARCH GO

COMPUTER HARDWARE
PCS & LAPTOPS
NETWORKING
ELECTRONICS
HOME THEATER
CAMERAS & CAMCORDERS
SOFTWARE
GAMING
CELL PHONES
HOME & OFFICE
MORE

Guided Search

Advanced Search

Power Search

Useful Links

New Product (35)

Discount Item (33)

Free Shipping (12)

Top Sellers (6)

Open Box (1)

Price

\$100 - \$200 (13)

\$200 - \$300 (17)

\$300 - \$400 (7)

Screen Size

17" (2)

19" (4)

20" (4)

21.5" (4)

22" (5)

24" (6)

Home > Categories > LCD Monitors (x) > Text Search Terms: samsung lcd (x) (1-20 of 37 Results)

Search Results

Related Searches: [samsung lcd monitor](#), [samsung lcd tv](#), [samsung led](#)

Search Within GO

Sort by: Featured Items View: 20

Select items COMPARE

Page 1 of 2 ◀ ▶ ↻

Compare

SAMSUNG 943BWT-1 Matte Black 19" 5ms w/height, Swivel, Tilt And Pivot Adjustable Stand Widescreen LCD Monitor

- + D-Sub: 1
- + DVI: 1
- + Built-in TV Tuner: No
- + Built-in Speakers: No
- + Model #: 943BWT-1
- + Item #: N82E16824001394
- + Return Policy: Monitor Replacement Only Return Policy

\$229.99

\$11.19 Shipping

ADD TO CART

Built-in digital TV tuner

SAMSUNG P2370HD Black 23" 5ms HDMI Widescreen LCD Monitor

- + D-Sub: 1

was: \$309.99

\$289.99

\$8.99 Shipping

◀ ▶ ↻ 2 / 77

MOTIVATING EXAMPLE

A search engine result list looks like this:

us action film site:en.wikipedia.org

Search

Web [+ Show options...](#)

Results 1 - 10 of about 43,900 from en.wikipedia.org for us action film. (0.20 seconds)



[Dragonball Evolution - Wikipedia, the free encyclopedia](#) ☆
 Dragonball Evolution is a 2009 American live-action film adaptation based upon the Japanese "U.S. Dragonball Film Site Launches with Updated Trailer". ...
[en.wikipedia.org/wiki/Dragonball_Evolution](#) - [Cached](#) - [Similar](#)



[Death Note \(film\) - Wikipedia, the free encyclopedia](#) ☆
 Live-action film. Death Note: The Last Name. Director, Shūsuke Kaneko "1st Death Note Film to Run in 300+ U.S. Theaters in May", Anime News Network. ...
[en.wikipedia.org/wiki/Death_Note_\(film\)](#) - [Cached](#) - [Similar](#)



[Missing in Action \(film\) - Wikipedia, the free encyclopedia](#) ☆
 Missing in Action is a 1984 action film directed by Joseph Zito and ... Colonel James Braddock (Chuck Norris) is a U.S. Army officer who spent seven years ...
[en.wikipedia.org/wiki/Missing_in_Action_\(film\)](#) - [Cached](#) - [Similar](#)




[Transformers \(film\) - Wikipedia, the free encyclopedia](#) ☆
 Transformers is a 2007 live-action film, based on the Transformers toy line. but leave the rest of us wondering if Hollywood could possibly aim lower" ...
[en.wikipedia.org/wiki/Transformers_\(film\)](#) - [Cached](#) - [Similar](#)



[Terminator \(franchise\) - Wikipedia, the free encyclopedia](#) ☆
 The American Film Institute (AFI) has also recognized both films on a number of ... The series premiere in the United States was watched by 18.6 million viewers ... "impressive and abundant action with realistic visual effects and, ...
[en.wikipedia.org/wiki/Terminator_\(franchise\)](#) - [Similar](#)

MOTIVATING EXAMPLE

In contrast, a faceted interface over the result list looks like this:


Facetedpedia


What kind of entities are you looking for:

Facets

Selected Categories:

<div style="background-color: #90EE90; padding: 2px; margin-bottom: 5px;"> Living people [4] </div> <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> [Will Smith] [4] [Bill Pullman] [1] [Jeff Goldblum] [1] [Dean Cain] [1] [Roland Emmerich] [1] </div> <div style="width: 45%;"> [Gene Hackman] [1] [Richard Matheson] [1] [Tony Scott] [1] [Bridget Moynahan] [1] </div> </div> <div style="background-color: #B0C4DE; padding: 2px; margin-bottom: 5px;"> Events by year [4] </div> <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> Establishments by year [4] Military history by year [2] Years in television [1] </div> <div style="width: 45%;"> Conflicts by year [3] Disestablishments by year [1] Introductions by year [1] </div> </div> <div style="background-color: #D2B48C; padding: 2px; margin-bottom: 5px;"> People by city in the United States by state [4] </div> <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> People by city in Pennsylvania [4] People by city in Illinois [2] People by city in Missouri [1] </div> <div style="width: 45%;"> People by city in New York [2] People by city in California [1] </div> </div> <div style="background-color: #90EE90; padding: 2px;"> California counties [4] </div>	<div style="margin-bottom: 10px;"> [remove] American film actors > [Will Smith] </div> <div style="background-color: #000080; color: white; padding: 5px; text-align: center;"> <h3>Wikipedia Articles</h3> </div> <div style="text-align: center; margin-bottom: 10px;"> • 4 Articles Selected </div> <div style="display: flex;"> <div style="width: 45%;"> <div style="background-color: #000080; color: white; padding: 2px; margin-bottom: 5px;"> Independence Day (film) </div> <p>The United States military originally intended to provide personnel, vehicles, and costumes for the film; however, they backed out when the producers ... http://en.wikipedia.org/wiki/Independence_Day_(film) </p> <div style="background-color: #000080; color: white; padding: 2px; margin-bottom: 5px;"> Enemy of the State (film) </div> <p>Language, English. Budget, US\$90000000. Gross revenue, US\$250649836 Janet Maslin of the New York Times approved of the film's action-packed sequences, ... http://en.wikipedia.org/wiki/Enemy_of_the_State_(film) </p> </div> <div style="width: 45%;"> <div style="background-color: #000080; color: white; padding: 2px; margin-bottom: 5px;"> I, Robot (film) </div> <p>I, Robot is a 2004 science-fiction action film directed by Alex Proyas. ... Produced with a budget of US\$120 million, the film grossed US\$144801023 ... http://en.wikipedia.org/wiki/I,_Robot_(film) </p> <div style="background-color: #000080; color: white; padding: 2px; margin-bottom: 5px;"> I Am Legend (film) </div> <p>I Am Legend was released on December 14, 2007, in the United States, and opened to and that there was "a pretty heavy screenplay for an action film. ... http://en.wikipedia.org </p> </div> </div>
--	---



FACETS VS. CLUSTERS

- Clusters are **partitions** of articles by topics
- Facets provides **multiple dimensional views** for a set of articles

[web](#) [news](#) [images](#) [wikipedia](#) [blogs](#) [jobs](#) [more »](#)

[advanced preferences](#)

Top 199 results of at least 1,446 retrieved for the query **us action film** ([details](#))

Search Results

clusters

remix

All Results (199)

- **Films** Category (41)
- **Fighter** (10)
- **Book** (10)
- **Anime** (7)
- **Death** (10)
- **Power** (10)
- **Lord** (7)
- **Martial arts** (5)
- **Record** (6)
- **Live-action feature film** (8)

- [Above the Law \(film\)](#)

Above the Law is a US 1988 **action film** directed by Andrew Davis in which a tough cop, Nico, plans to take down the CIA. Nico family has Mafia contacts. Nico is a 6th dan aikido black belt, a Vietnam ... There is a rap group called Above the Law . Category:1988 **films** Category: **Films** directed by Andrew Davis [en.wikipedia.org/wiki/Above_the_Law_\(film\)](#)
- [Danny the Dog \(film\)](#)


Danny the Dog , released in the US, UK and Australia as Unleashed , is a 2005 **action / thriller film** , directed by Louis Leterrier .It is set in Glasgow , Scotland .Plot summaryDanny (Jet Li) is a Chinese slave with the mind and personality of a young child. Forced to wear a dog ... External links Official site Category:2005 **films** Category:Martial arts **films** Category:Thriller **films** Unleashed DVD Official Universal Studios Site [en.wikipedia.org/wiki/Danny_the_Dog_\(film\)](#)
- [Street Fighter II](#)


... Street Fighter . The animated **movie** is more popular with fans than the live **action**

5 / 77

OUR GOAL - FACETEDPEDIA

- A system generates **faceted interfaces** for a set of Wikipedia articles (target articles)
- Facets are **query-dependent**
 - Query="us action film", Facets=actors, directors, film production companies,...
 - Query="Texas university", Facets=cities in Texas, alumni, football teams, ...
- Generation is **automatic** compared with manual or predefined approach

OUR GOAL - FACETEDPEDIA

- A system generates **faceted interfaces** for a set of Wikipedia articles (target articles)
- Facets are **query-dependent**
 - Query="us action film", Facets=actors, directors, film production companies,...
 - Query="Texas university", Facets=cities in Texas, alumni, football teams, ...
- Generation is **automatic** compared with manual or predefined approach

OUR GOAL - FACETEDPEDIA

- A system generates **faceted interfaces** for a set of Wikipedia articles (target articles)
- Facets are **query-dependent**
 - Query="us action film", Facets=actors, directors, film production companies,...
 - Query="Texas university", Facets=cities in Texas, alumni, football teams, ...
- Generation is **automatic** compared with manual or predefined approach

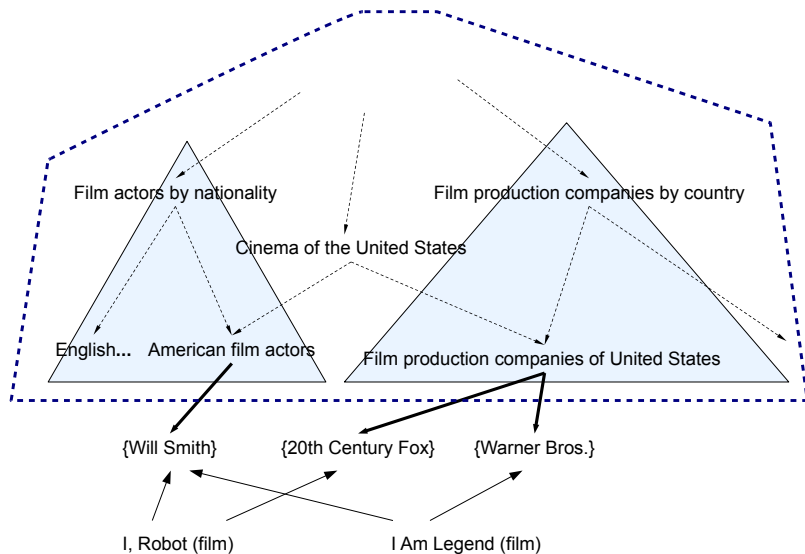
OUR GOAL - FACETEDPEDIA

- A system generates **faceted interfaces** for a set of Wikipedia articles (target articles)
- Facets are **query-dependent**
 - Query="us action film", Facets=actors, directors, film production companies,...
 - Query="Texas university", Facets=cities in Texas, alumni, football teams, ...
- Generation is **automatic** compared with manual or predefined approach

OUR GOAL - FACETEDPEDIA

- A system generates **faceted interfaces** for a set of Wikipedia articles (target articles)
- Facets are **query-dependent**
 - Query="us action film", Facets=actors, directors, film production companies,...
 - Query="Texas university", Facets=cities in Texas, alumni, football teams, ...
- Generation is **automatic** compared with manual or predefined approach

HOW TO?



CHALLENGE 1

What are our facets?

← dimensions formed by **attribute articles** of target articles

CHALLENGE 1

What are our facets?

← dimensions formed by **attribute articles** of target articles

I Am Legend (film)

From Wikipedia, the free **target article**

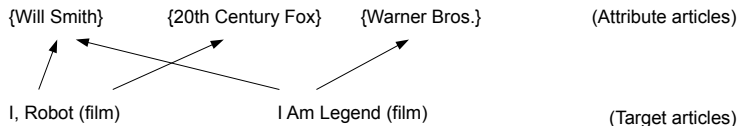
I Am Legend is a 2007 film directed by Francis Lawrence and starring **Will Smith**. It is the third feature film adaptation of Richard Matheson's 1954 novel of the same name, following 1964's *The Last Man on Earth* and 1971's *The Omega Man*.^[2] Smith plays **virulent** Robert Neville, who is immune to a vicious man-made **attribute articles**. He works to create a refuge where living in Manhattan in 2012, a city inhabited by violent victims of the virus. The film's plot is an example of a **zombie apocalypse** story.

Warner Bros. began developing *I Am Legend* in 1994, and various actors and directors were attached to the project, though production was delayed due to budgetary concerns related to the script. Production began in 2006 in New York

CHALLENGE 1

What are our facets?

⇐ dimensions formed by **attribute articles** of target articles



CHALLENGE 1

How to group attribute articles?

← by their categories

CHALLENGE 1

How to group attribute articles? ⇐ by their categories

Will Smith **attribute article**

From Wikipedia, the free encyclopedia

"Fresh Prince" redirects here. For the TV series, see The Fresh Prince of Bel-Air.

"W.Smith" redirects here. For the British retailer, see WH Smith. For other persons with similar names, see William Smith.

Willard Christopher "Will" Smith, Jr. (born September 25, 1968)^[1] is an American actor, film producer and rapper. He has enjoyed success in music, television and film. In April 2007, *Newsweek* called him the most powerful actor on the planet.^[2] Smith has been nominated for four *Golden Globe Awards*, two *Academy Awards*, and has won multiple *Grammy Awards*.

In the late 1980s, Smith achieved modest fame as a rapper under the name **The Fresh Prince**. In 1990, his celebrity increased dramatically when he starred in the popular television series *The Fresh Prince of Bel-Air*. The show ran for nearly six years (1990–1996) on *NBC* and has been syndicated consistently on various networks since then. In the mid-1990s, Smith transitioned from television to film, and ultimately starred in numerous **blockbuster films** that received broad box office success. In fact, he is the only actor in history to have eight consecutive films gross over \$100 million in the domestic box office as well as being the only actor to have eight consecutive films in which he starred open at the #1 spot in the domestic box office tally.^[3] Fourteen of the 19 fiction films he has acted in have accumulated a worldwide grosses of over \$100 million, and 4 of them took in over \$500 million in global box office receipts. His most financially successful films have been *Bad Boys*, *Bad Boys II*, *Independence Day*, *Men in Black II*, *I, Robot*, *The Pursuit of Happyness*, *I Am Legend*, *Hancock*, *Wild Wild West*, *Enemy of the State*, *Shark Tale*, *Hitch* and *Seven Pounds*. He also earned critical praise for his performances in *Six Degrees of Separation*, *Ali* and *The Pursuit of Happyness*, receiving Best Actor Oscar nominations for the latter two.

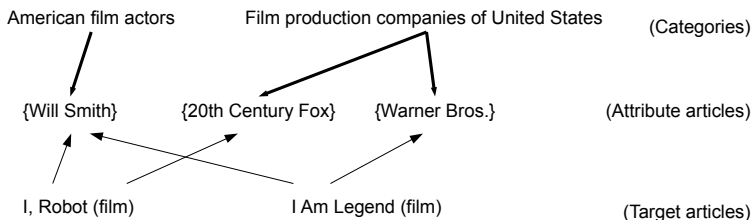
categories

Categories: 1968 births | African-American people | African American actors | American Christians | American film actors | American hip hop musicians | American rappers | American television actors | Columbia Records artists | East Coast rappers | Grammy Award winners | Interscope Records artists |

CHALLENGE 1

How to group attribute articles?

⇐ by their categories



CHALLENGE 1

How to organize categories?

← by category hierarchy

CHALLENGE 1

How to organize categories?

⇐ by category hierarchy

Category: American film actors

category

From Wikipedia, the free encyclopedia

Subcategories

This category has the following 2 subcategories, out of 3 total.

*

- [\[+\] African American film actors \(157 P\)](#)
- [\[+\] American silent film actors \(1556 P\)](#)

subcategories

Pages in category "American film actors"

The following 198 pages are in this category, out of 13,301 total. This list may not reflect recent changes ([learn more](#)).

A

- Beverly Aadland
- Mariann Aalda
- Aaliyah
- Angela Aames
- Willie Aames
- Caroline Aaron
- Quinton Aaron
- Victor Aaron
- Peter Abbay
- Dorothy Abbott

A cont.

- Kathryn Adams Doty
- Thom Adcox-Hernandez
- Calpernia Addams
- Jane Adams (actress)
- Nancy Addison
- Wesley Addy
- Traci Adell
- Jason Adelman
- Paul Adelstein
- Alexandra Adi

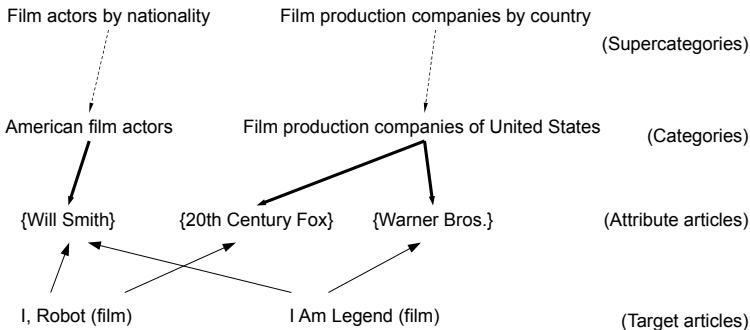
Categories: Cinema of the United States | American actors by medium | Film actors by nationality

supercategories

CHALLENGE 1

How to organize categories?

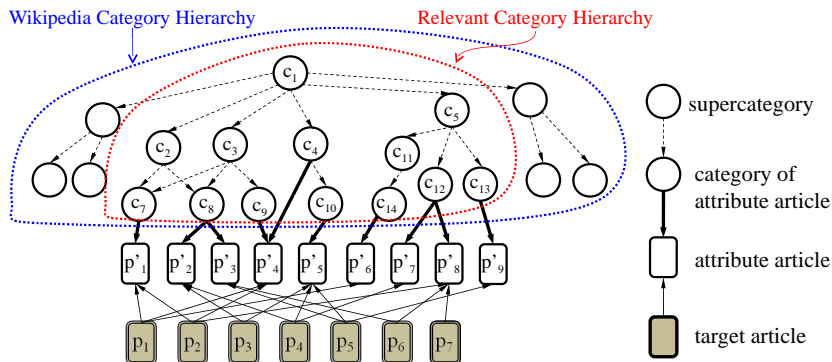
⇐ by category hierarchy



THE CONCEPT OF FACET

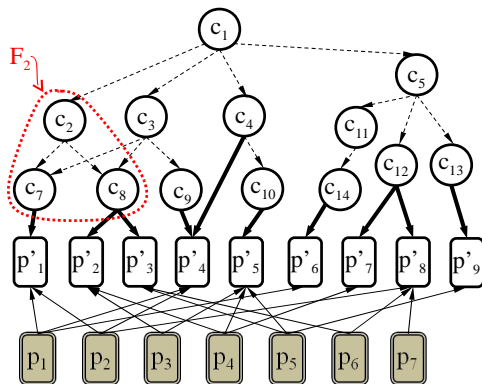
establish the concept of facet by **"structures"** from Wikipedia

- **RCH**: Relevant Category Hierarchy



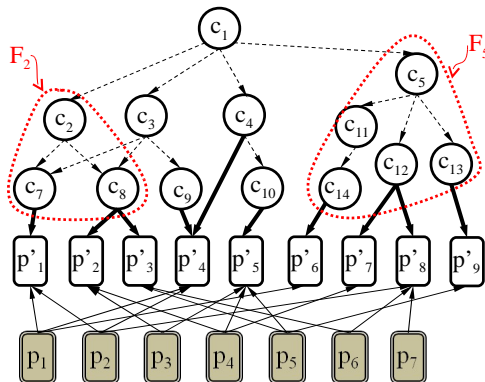
THE CONCEPT OF FACET - CONT.

- **facet**: an induced subgraph of a RCH, e.g., F_2 is a facet with $\{c_2, c_7, c_8\}$



THE CONCEPT OF FACET - CONT.

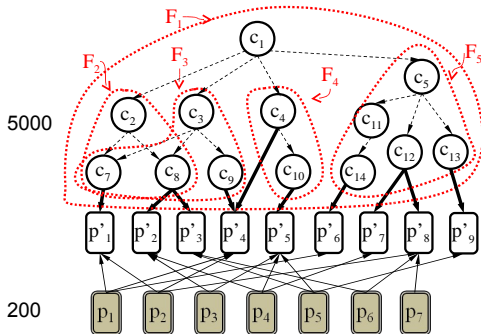
- faceted interface**: an interface consists of k facets, e.g., $I = \{F_2, F_5\}$ is a faceted interface with 2 facets ($k=2$)



CHALLENGE 2

- 1 What are our facets?
- 2 How to pick good facets and faceted interfaces?

Single-Facet Ranking & Multi-Facet Ranking



SINGLE-FACET RANKING

Intuitions:

- 1 How much user effort is needed when navigating through a single facet.
- 2 How many target articles a single facet could reach or cover.

SINGLE-FACET RANKING

Intuitions:

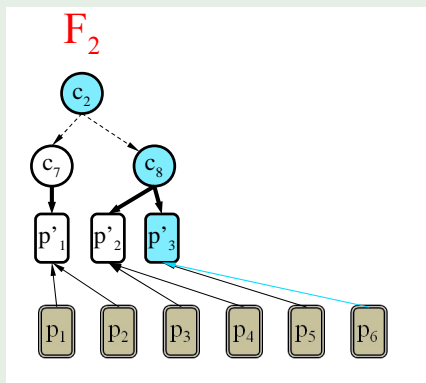
- 1 How much user effort is needed when navigating through a single facet.
- 2 How many target articles a single facet could reach or cover.

COST OF NAVIGATIONAL PATH

capture first intuition (user effort) by cost of **navigational path**:

EXAMPLE

navigational path: $L = c_2 \dashrightarrow c_8 \Rightarrow p'_3 \leftarrow p_6$

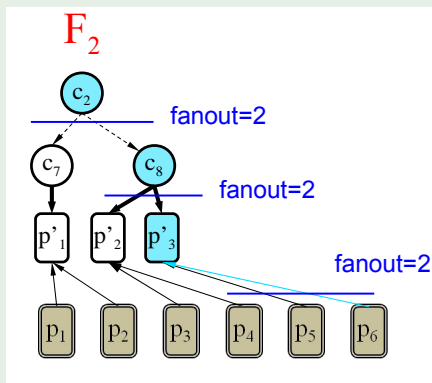


COST OF NAVIGATIONAL PATH

- **cost of navigational path**: sum of logarithmic fanout

EXAMPLE

$$\begin{aligned} \text{cost}(L) &= \sum_{\text{step}=c_2, c_8, p'_3} \log_2(\text{fanout}(\text{step})) \\ &= \log_2(2) + \log_2(2) + \log_2(2) = 3. \end{aligned}$$

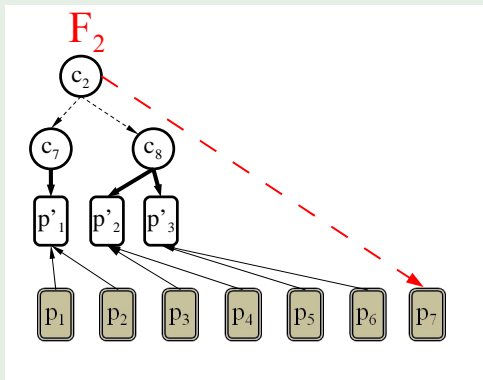


PENALTY PATH

capture second intuition (coverage) by adding a high cost pseudo **penalty path** for unreachable targeted articles:

EXAMPLE

penalty path: $L' = c_2 \dashrightarrow p_7$

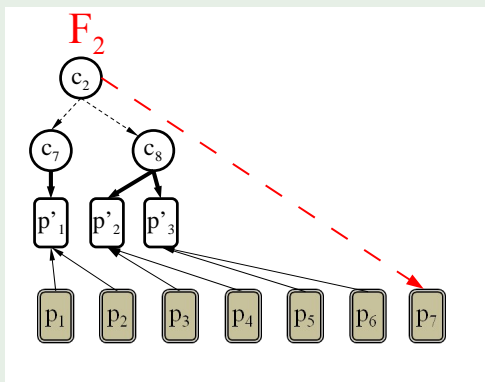


COST OF SINGLE FACET

- **cost of facet**: average cost of navigational paths + **penalty**
(**penalty** value is determined empirically)

EXAMPLE

$$\text{cost}(F_2) = \frac{1}{7} \times (\sum \text{cost}(\text{path } L_{p \in \{p_1, \dots, p_6\}}) + \text{penalty} \times \#L'_{p \in \{p_7\}})$$



MULTI-FACET RANKING

Intuitions:

- 1 overall cost of k facets should be small

MULTI-FACET RANKING

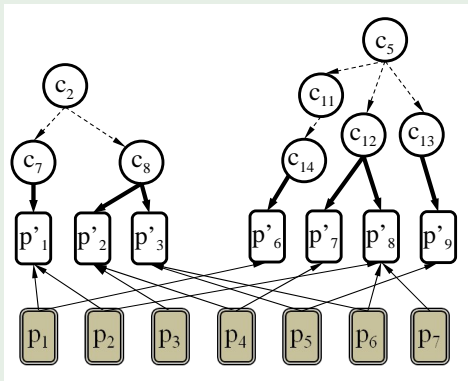
Intuitions:

- 1 overall cost of k facets should be small
- 2 k facets should NOT **overlap** a lot

USER NAVIGATION MODEL

EXAMPLE

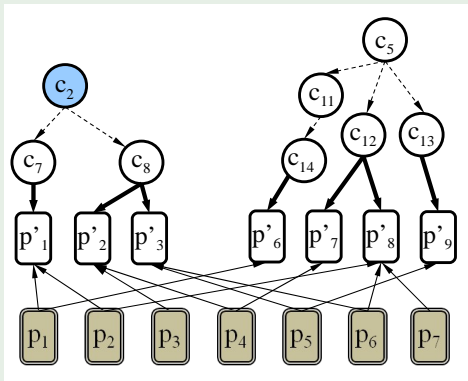
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

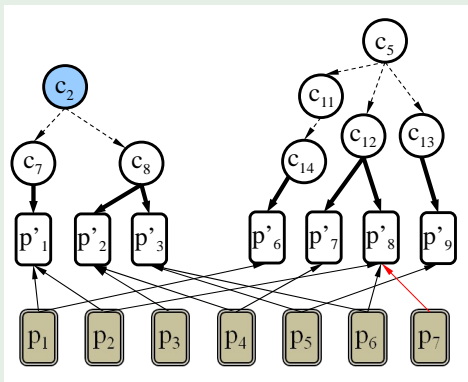
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

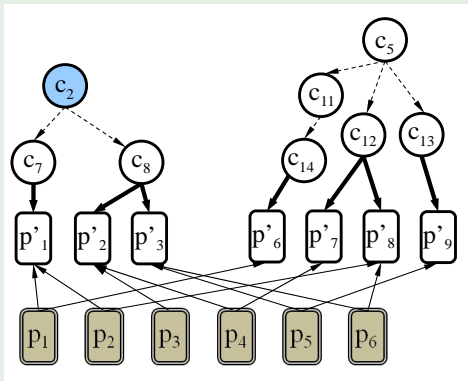
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

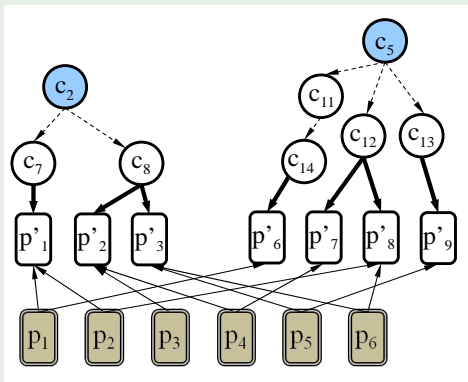
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

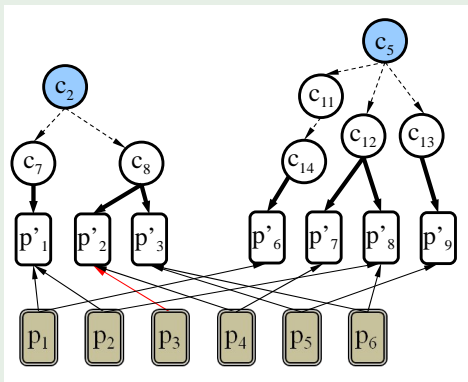
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

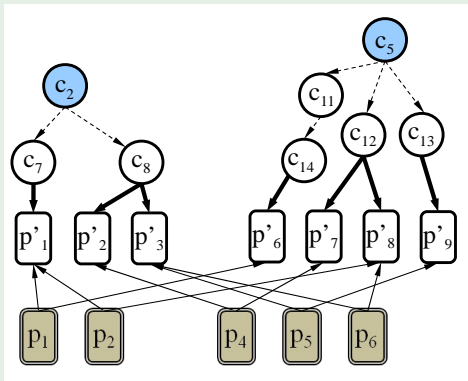
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

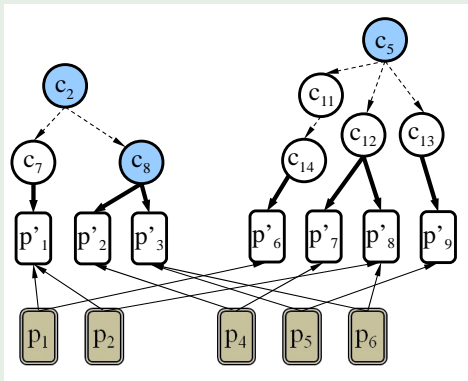
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

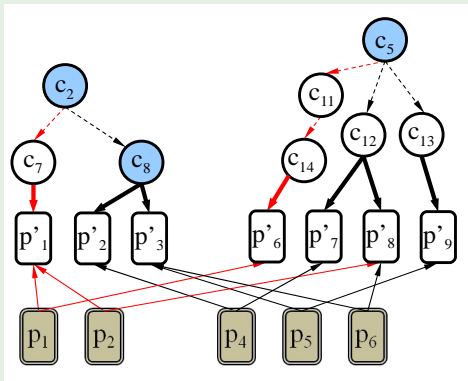
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

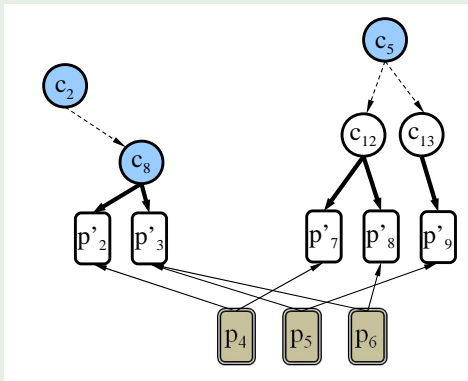
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

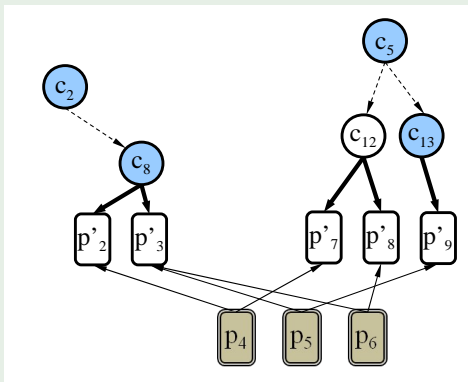
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

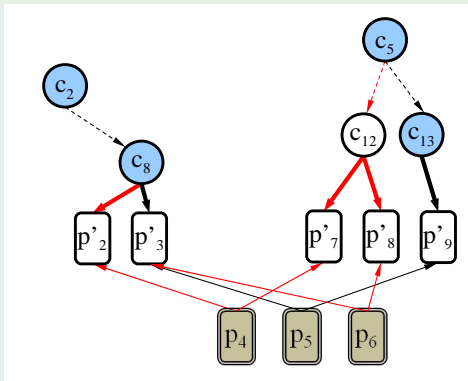
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

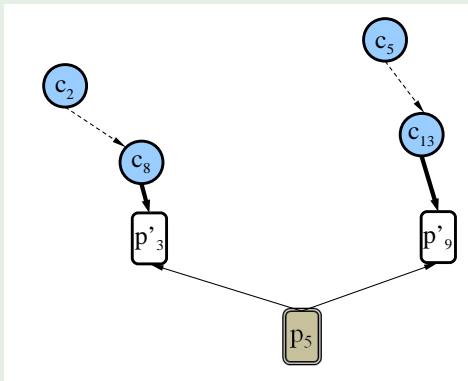
Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...



USER NAVIGATION MODEL

EXAMPLE

Given a faceted interface with 2 non-overlapping facets
 $I = \{F_2, F_5\}$, a user navigates like...

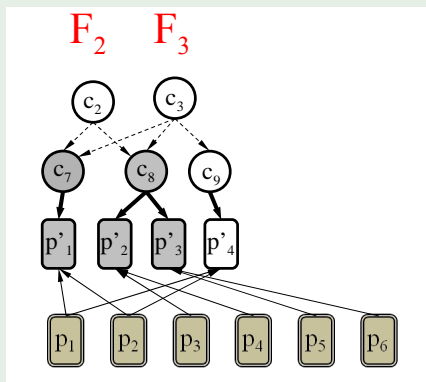


MULTI-FACET RANKING

capture second intuition (overlap) by **Jaccard similarity** between 2 facets:

EXAMPLE

$$\text{sim}(F_2, F_3) = \frac{|\{c_7, c_8\}| + |\{p'_1, p'_2, p'_3\}|}{|\{c_2, c_7, c_8, c_3, c_9\}| + |\{p'_1, p'_2, p'_3, p'_4\}|} = 5/9.$$

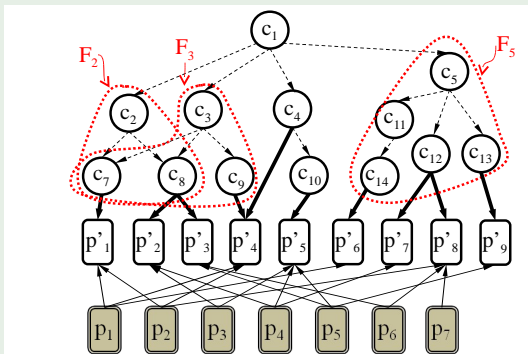


MULTI-FACET RANKING

and by **average pair-wise similarity**(AVG) between k facets, e.g., a faceted interface with 3 facets $\{F_2, F_3, F_5\}$:

EXAMPLE

$$AVG(\{F_2, F_3, F_5\}) = \frac{1}{3} \times (sim(F_2, F_3) + sim(F_2, F_5) + sim(F_3, F_5))$$



CHALLENGE 3

- 1 What are our facets?
- 2 How to pick good facets and faceted interfaces?
- 3 How to do the rankings efficiently for Wikipedia?

Facet Discovery Algorithm

FACET DISCOVERY ALGORITHM

Step 1: Retrieve attribute articles & Build RCH

Input: target articles

Output: attribute articles, RCH

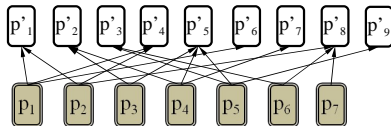


FACET DISCOVERY ALGORITHM

Step 1: Retrieve attribute articles & Build RCH

Input: target articles

Output: attribute articles, RCH

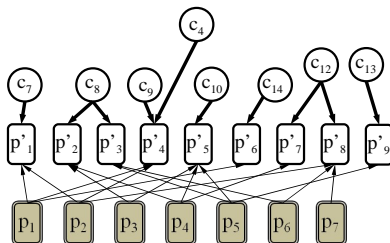


FACET DISCOVERY ALGORITHM

Step 1: Retrieve attribute articles & Build RCH

Input: target articles

Output: attribute articles, RCH

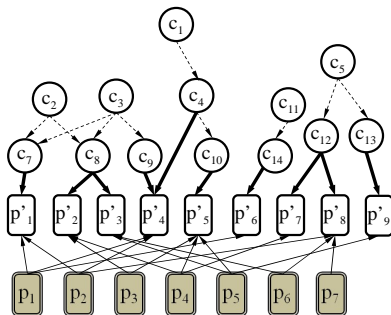


FACET DISCOVERY ALGORITHM

Step 1: Retrieve attribute articles & Build RCH

Input: target articles

Output: attribute articles, RCH

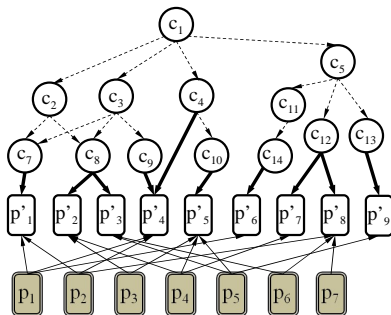


FACET DISCOVERY ALGORITHM

Step 1: Retrieve attribute articles & Build RCH

Input: target articles

Output: attribute articles, RCH

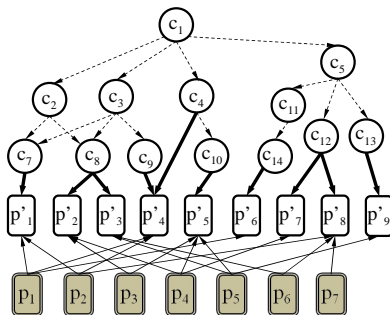


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs

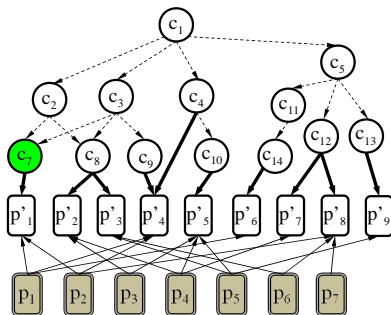


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs

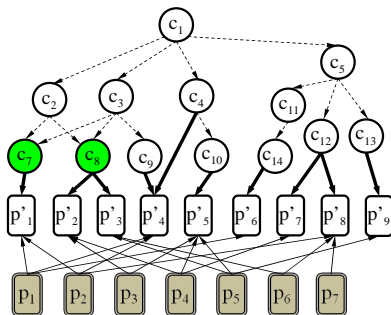


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs

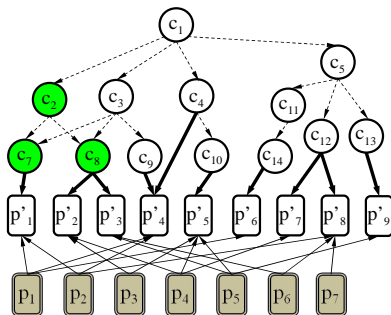


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs

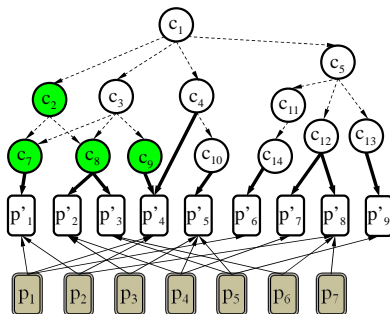


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs

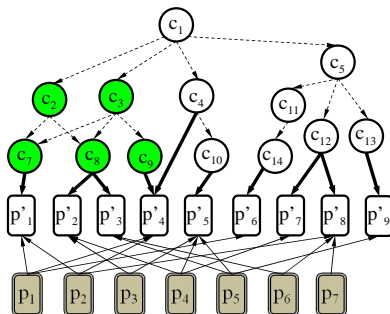


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs

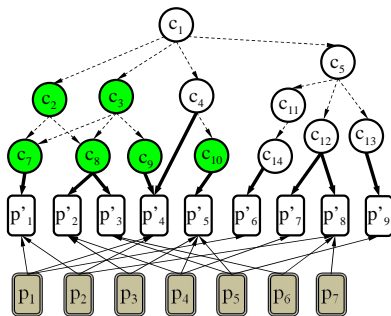


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs

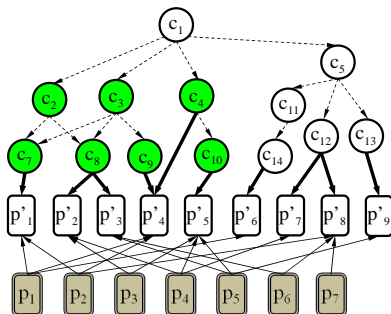


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs

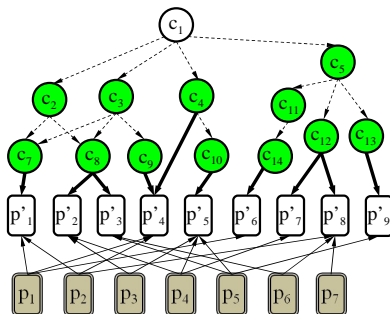


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs

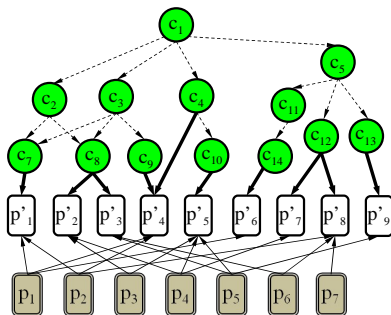


FACET DISCOVERY ALGORITHM

Step 2: Rank single facets based on their cost

Input: target articles, attribute articles, RCH

Output: a ranked list of single facets and their costs



FACET DISCOVERY ALGORITHM

Step 3: Select multi-facet using hill-climbing method

Input: top- n facets from the ranked list and their costs C

Output: a faceted interface with k facets

- 1 randomly pick k facets from n facets
- 2 compute $cost$, AVG of k facets
- 3 **repeat**
 - switch facet i in k facets with facet j from remaining
 - compute $cost_{new}$
 - compute AVG_{new}
 - **if** ($cost_{new} \downarrow$ && $AVG_{new} \downarrow$) **then** replace facet i with j**until** k facets are stable
- 4 **return** k facets

RELATED WORKS

semantics of hierarchies rich shallow	Endeca Mercado [Ben-Yitzhak08] [Debabrata08] [Yee03]	Facetedpedia Castanet
	FacetedDBLP [Dakka05,Dakka08] [Roy08]	
	structured	text

types of facets

hierarchy construction automatically discovered predefined	[Dakka08] [Dakka05]	Facetedpedia Castanet
	FacetedDBLP [Ben-Yitzhak08] [Yee03] [Debabrata08] Endeca Mercado [Roy08]	
	predefined	automatically discovered

facet identification

Hierarchy

shallow: People→Will Smith

rich: People→Actor→Film Actor→American Film Actor→Will Smith

RELATED WORKS

Castanet [StoicaHearst07]

Recipes Powered by Flamenco

Pine Save Search History and Settings Return to Search New Search Logout

Username Password

[Create a New Account](#)

Show tooltip previews of subcategories

FLAVORER		DISH	
almond_extract (40)	condiment (1912)	applesauce (28)	burrito (2)
bouillon_cube (14)	coriander (3)	apple_sauce (2)	carmeloni (11)
caraway_seed (12)	corn_powder (45)	barbecue (43)	cariot_pudding (3)
cayenne_pepper (62)	dill_seed (4)	barbecued_spareribs (3)	casserole (466)
celery_salt (19)	garlic (497)	beef_stroganoff (7)	cheese_fondue (2)
celery_seed (27)	more...	boiled_egg (9)	more...
chili_powder (73)		buffalo_wing (3)	
VEGETABLE		MEAL	
artichoke (14)	cauliflower (25)	banquet (3)	picnic (13)
artichoke_heart (10)	celery (306)	breakfast (54)	pottuck (5)
asparagus (13)	cucumber (25)	dinner (43)	snack (41)
beet (10)	eggplant (20)	lunch (19)	supper (12)
broccoli (49)	greens (556)	luncheon (2)	

- not for general articles
 - short
 - domain-specific (e.g. recipes...)
 - limited vocabulary
- hierarchy is shallow

RELATED WORKS

Faceted Wikipedia Search [http://dbpedia.neofonie.de/browse/]



search powered by neofonie

[About neofonie](#)[About DBpedia](#)[Imprint](#)[Help](#)[First](#) | [Previous](#) | [Next](#) | [Last](#)

▼ item type

Work (77)

Film (77)

Musical Work (6)

[more](#)

▼ language

English language (37)

Standard Cantonese (7)

Thai language (2)

[more](#)

▼ released in year

2008 (6)

2007 (6)

2000 (5)

[more](#)

▼ runtime (minutes)

01:000 (5)

Your Filters

[Reset Filters](#) ✕

Results 1 to 6 of 77

text search for us action film ✕ item type Film ✕

Transformers (film)

Transformers is a 2007 live-action film adaptation of the Transformers franchise, directed by Michael Bay and written by Roberto Orci and Alex Kurtzman. It stars Shia LaBeouf as Sam Witwicky, a teenager involved in a war between the heroic Autobots and the evil Decepticons, two factions of alien robots who can disguise themselves by transforming into everyday machinery.

Predator (film)

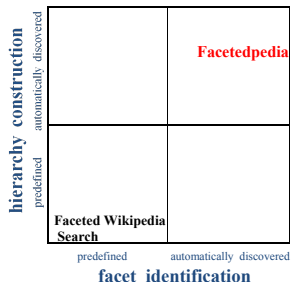
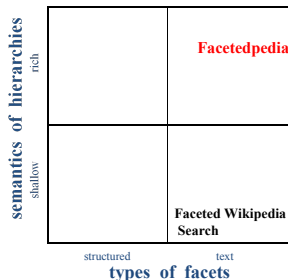
Predator is a 1987 science fiction, action and horror film directed by John McTiernan, starring Arnold Schwarzenegger, Carl Weathers, Jesse Ventura, and Kevin Peter Hall. The story follows an elite team on a mission to rescue hostages from a guerrilla group in Central America. Unknowingly, the group is hunted by an extraterrestrial lifeform. Reaction to the film was generally favorable, and the film grossed US\$60 million in the United States.

Crouching Tiger, Hidden Dragon

Crouching Tiger, Hidden Dragon is a Chinese-language film in the wuxia style, released in 2000. A China-Hong Kong-Taiwan-United States co-production, the film was directed by Ang Lee and featured an international cast of ethnic Chinese actors, including Chow Yun-Fat, Michelle Yeoh, Zhang Ziyi and Chang Chen. The movie was based on the fourth novel in a pentology, known in China as the Green Iron Pentology, by wuxia novelist Wang Duxi.

RELATED WORKS

Faceted Wikipedia Search



- facets:
 - query independent for specific domain
 - probably from infobox attributes, similar to relational data, no facet identification
 - only for articles with infobox
- no hierarchy

EXPERIMENT SETTINGS

- Raw Data: Wikipedia dump 20080724, 2.4M articles, 110M hyperlinks, 330K categories, 730K category links in category hierarchy
- Programs: Preprocessing of Wikipedia dump (removing irrelevant articles and categories, removing cycles in category hierarchy)
- Parameters: #target articles=200, #facets=10

A USER STUDY

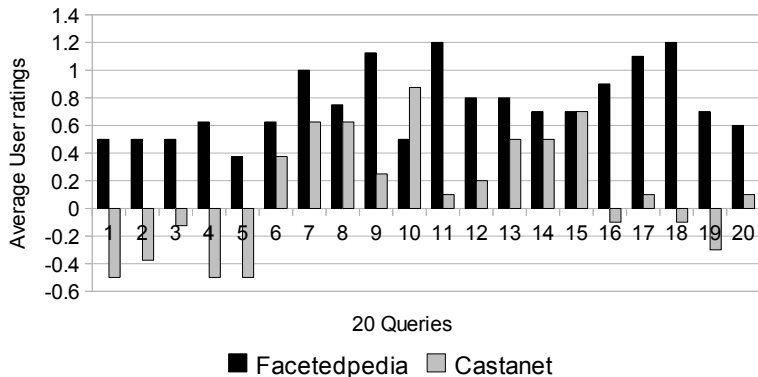
- select 20 queries from different domains
- generate faceted interfaces for each query by **Facetedpedia** and **Castanet**
- each interface has been explored and evaluated by 8~10 user

Q1	action film	Q2	country singer
Q3	philosophers	Q4	Texas universities
Q5	Turing Award winner	Q6	missile
Q7	Ivy League schools	Q8	NBA players
Q9	historic landmarks	Q10	cartoon characters
Q11	Microsoft acquired game companies	Q12	stand up comedian
Q13	graph theorists	Q14	lakes in North America
Q15	American presidents	Q16	battle far east
Q17	waterfall national park	Q18	Chinese cuisine
Q19	premier league clubs	Q20	PS3 game

A USER STUDY - CONT.

Average user ratings of Facetedpedia vs. Castanet

(Rating scale: 2 Very useful; 1 Useful; 0 Useful to some extent; -1 Not Very Useful; -2 Useless)



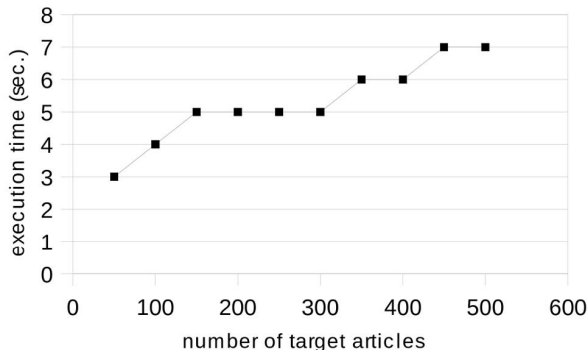
QUALITY EVALUATION

Measurement of coverage, average width, depth, pairwise similarity of faceted interfaces.

	Coverage	average width	average height	average pair-wise similarity
Random-k	72.3%	53.8	8.6	0.108
Top-k	73.9%	10.2	5.5	0.187
Hill-climbing	68.9%	9.8	5.7	0.072

EFFICIENCY EVALUATION

Run time varies for different number of target articles almost linearly.



CONCLUSIONS

- 1 Facetedpedia System
- 2 Facet Ranking Metrics
- 3 Faceted Interface Discovery Algorithm
- 4 Evaluation