Utilizing Query Facets for Search Result Navigation

Jan Friedrich, Christoph Lindemann, Michael Petrifke

Computer Networks and Distributed Systems
University of Leipzig

01.09.2015

UNIVERSITÄT
LEIPZIG
1409
2009
ALMA MATER
LIPSIENSIS

Table of contents

Introduction to Faceted Web Search

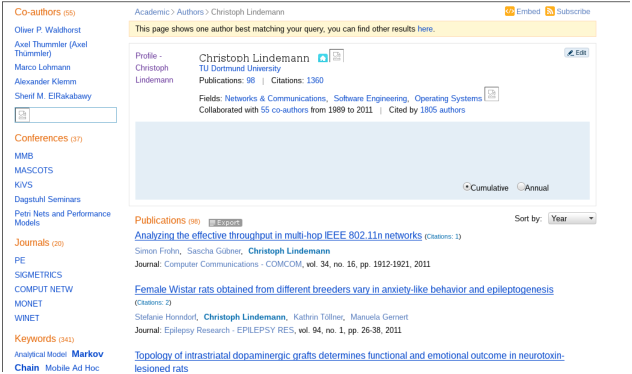
Facet Generation

Navigational Properties

Facet Feedback

Evaluation

Examples I: Microsoft Academic Search



Co-authors (55)

Oliver P. Waldhorst

Axel Thummler (Axel Thümmler)

Marco Lohmann

Alexander Klemm

Sherif M. ElRakabawy

Conferences (37)

MMB

MASCOTS

KiVS

Dagstuhl Seminars

Petri Nets and Performance Models

Journals (20)

PE

SIGMETRICS

COMPUT NETW

MONET

WINET

Keywords (341)

Analytical Model **Markov Chain** Mobile Ad Hoc

Academic > Authors > Christoph Lindemann

Embed Subscribe

This page shows one author best matching your query, you can find other results here.

Profile - Christoph Lindemann

Christoph Lindemann

Edit

TU Dortmund University

Publications: 98 | Citations: 1360

Fields: Networks & Communications, Software Engineering, Operating Systems

Collaborated with 55 co-authors from 1989 to 2011 | Cited by 1805 authors

● Cumulative ○ Annual

Publications (98) Export

Sort by: Year

Analyzing the effective throughput in multi-hop IEEE 802.11n networks (Citations: 1)

Simon Frohn, Sascha Gübner, **Christoph Lindemann**

Journal: Computer Communications - COMCOM, vol. 34, no. 16, pp. 1912-1921, 2011

Female Wistar rats obtained from different breeders vary in anxiety-like behavior and epileptogenesis (Citations: 2)

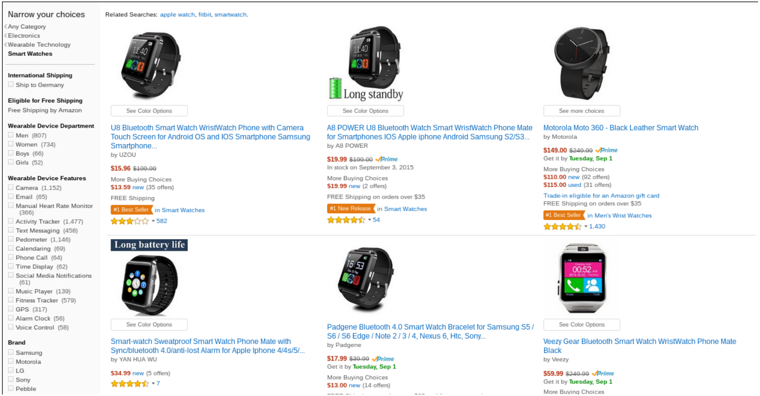
Stefanie Honndorf, **Christoph Lindemann**, Kathrin Töllner, Manuela Gernert

Journal: Epilepsy Research - EPILEPSY RES, vol. 94, no. 1, pp. 26-38, 2011

Topology of intrastriatal dopaminergic grafts determines functional and emotional outcome in neurotoxin-lesioned rats

Figure: Example of facets on Microsoft Academic Search

Examples II: Amazon.com



Narrow your choices

‹ Any Category
‹ Electronics
‹ Wearable Technology
Smart Watches

International Shipping

- □ Ship to Germany

Eligible for Free Shipping

Free Shipping by Amazon

Wearable Device Department

- □ Men (807)
- □ Women (734)
- □ Boys (66)
- □ Girls (52)

Wearable Device Features

- □ Camera (1,152)
- □ Email (65)
- □ Manual Heart Rate Monitor (366)
- □ Activity Tracker (1,477)
- □ Text Messaging (458)
- □ Pedometer (1,146)
- □ Calendaring (69)
- □ Phone Call (64)
- □ Time Display (62)
- □ Social Media Notifications (61)
- □ Music Player (139)
- □ Fitness Tracker (579)
- □ GPS (317)
- □ Alarm Clock (56)
- □ Voice Control (58)

Brand

- □ Samsung
- □ Motorola
- □ LG
- □ Sony
- □ Pebble

Related Searches: apple watch, fitbit, smartwatch.

U8 Bluetooth Smart Watch WristWatch Phone with Camera Touch Screen for Android OS and IOS Smartphone Samsung Smartphone...
by UZOU
$15.96 ~~$100.00~~
More Buying Choices
$13.59 new (35 offers)
FREE Shipping
#1 Best Seller in Smart Watches
★★★☆☆ 582

A8 POWER U8 Bluetooth Watch Smart WristWatch Phone Mate for Smartphones IOS Apple iphone Android Samsung S2/S3...
by A8 POWER
$19.99 ~~$100.00~~ ✓Prime
In stock on September 3, 2015
More Buying Choices
$19.99 new (2 offers)
FREE Shipping on orders over $35
#1 New Release in Smart Watches
★★★★☆ 54

Motorola Moto 360 - Black Leather Smart Watch
by Motorola
$149.00 ~~$249.99~~ ✓Prime
Get it by **Tuesday, Sep 1**
More Buying Choices
$110.00 new (92 offers)
$115.00 used (31 offers)
Trade-in eligible for an Amazon gift card
FREE Shipping on orders over $35
#1 Best Seller in Men's Wrist Watches
★★★★☆ 1,430

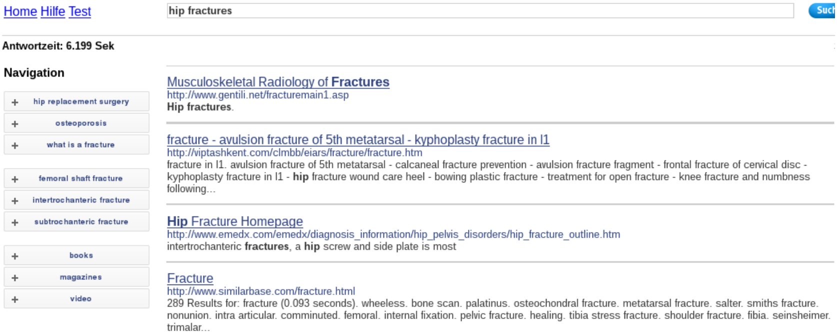
Smart-watch Sweatproof Smart Watch Phone Mate with Sync/bluetooth 4.0/anti-lost Alarm for Apple Iphone 4/4s/5/...
by YAN HUA WU
$34.99 new (5 offers)
★★★★☆ 7

Padgene Bluetooth 4.0 Smart Watch Bracelet for Samsung S5 / S6 / S6 Edge / Note 2 / 3 / 4, Nexus 6, Htc, Sony...
by Padgene
$17.99 ~~$39.99~~ ✓Prime
Get it by **Tuesday, Sep 1**
More Buying Choices
$13.00 new (14 offers)

Veezy Gear Bluetooth Smart Watch WristWatch Phone Mate Black
by Veezy
$59.99 ~~$249.99~~ ✓Prime
Get it by **Tuesday, Sep 1**
More Buying Choices

Figure: Example of facets on Amazon.com

Examples III: Implemented Results



Antwortzeit: 6.199 Sek

Musculoskeletal Radiology of **Fractures**
http://www.gentili.net/fracturemain1.asp
Hip fractures.

fracture - avulsion fracture of 5th metatarsal - kyphoplasty fracture in l1
http://viptashkent.com/clmbb/eiars/fracture/fracture.htm
fracture in l1. avulsion fracture of 5th metatarsal - calcaneal fracture prevention - avulsion fracture fragment - frontal fracture of cervical disc - kyphoplasty fracture in l1 - **hip** fracture wound care heel - bowing plastic fracture - treatment for open fracture - knee fracture and numbness following...

Hip Fracture Homepage
http://www.emedx.com/emedx/diagnosis_information/hip_pelvis_disorders/hip_fracture_outline.htm
intertrochanteric **fractures**, a **hip** screw and side plate is most

Fracture
http://www.similarbase.com/fracture.html
289 Results for: fracture (0.093 seconds). wheeless. bone scan. palatinus. osteochondral fracture. metatarsal fracture. salter. smiths fracture. nonunion. intra articular. comminuted. femoral. internal fixation. pelvic fracture. healing. tibia stress fracture. shoulder fracture. fibia. seinsheimer. trimalar...

Figure: Facets generated for the query "hip fractures"

Introduction to Facets

- A facet is a flat set of terms
- Facets provide selectors / filters for, mostly nominal, object attributes
- Displayed facets traditionally represent existing attributes of the listed objects
- Facets show aspects that help to easily distinguish objects on the level of one attribute → it is of no use to show a specific facet if all relevant objects match one and the same value of the corresponding attribute
- Facets provide insight and help to navigate the search result space

Faceted Web Search Characteristics

- Semi-structured documents
- Some explicit document attributes like in document reference systems (e.g. author, title, publication date, keywords) → however, not useful in the context of general web search
- Useful facets are not connected to predefined document attributes (e.g. search results for "IFA Berlin" might benefit from the facets "vendors" or "exhibition hall" → this information is hidden in the text)
- Huge number of possible facets and facet terms → every existing taxonomy provides many sets of related terms

Requirements of Faceted Web Search Systems

- Behave similar to Boolean filters → learned behavior from other faceted applications
- Terms of one and the same facet should be mutual exclusive → only few terms match the same document
- Small number of facets and terms per facet → facets distract the user
- Proposal: Use ranking features that characterize the partition properties of the candidate facets

Faceted Web Search Problems

1. Generation of facets and assignment of facet terms to documents
2. Ranking and selection of relevant facets for the user and query
3. Utilization of user-selected facet terms (user feedback)

First Work on Facet Generation

- Facetedpedia: Wikipedia provides categories and hyperlinks between articles [5]
- Blogs provide keywords and categories
- External resources like WordNet's hypernym information [1] and other taxonomies
- Above methods not applicable to the general web or require expensive offline computations
- Topic discovery, search result clustering → search for labels that fit subsets of the result documents → facet generation searches for one-level hierarchies that are representative for the search results

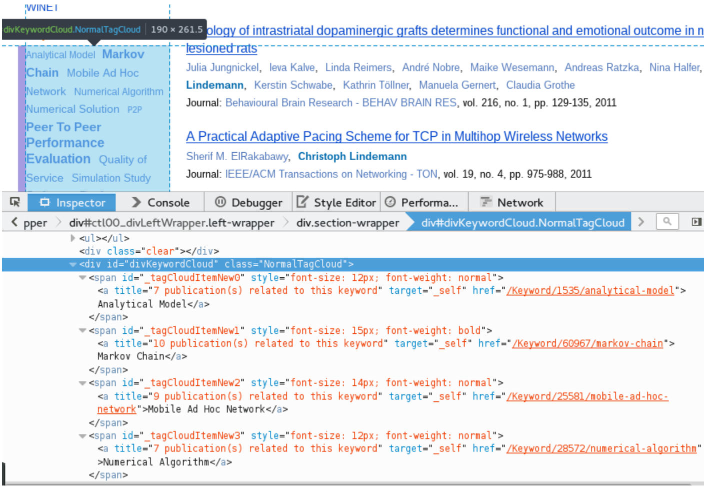
Facet Extraction from Lists

- Dou et al. [2] introduced the idea to exclusively utilize lists of terms that can be found in the search result documents → no external resources required
- Types of lists:
  - Lists in free text
  - Fixed HTML patterns (e.g. ol, ul and tables)
  - Visual repeat regions to extract lists that use CSS and other HTML structures than the fixed patterns above
- Above lists (list candidates) are post-processed, clustered and than ranked to generate the final facets

HTML Meta Patterns

- Modern web design sometimes utilizes Cascading Style Sheets (CSS) to generate visual lists from general HTML tags like span or p
- Observation I: fixed HTML patterns are not able to extract these lists
- Observation II: visual information is not required to extract most of these lists
- Proposal: HTML Meta Pattern, that finds elements whose children are mostly structurally identical (i.e. same HTML subtree based on the element names)
- ignore comments, script, ...

HTML Meta Pattern Example


WINET
divKeywordCloud.NormalTagCloud | 190 × 261.5
Analytical Model Markov Chain Mobile Ad Hoc Network Numerical Algorithm Numerical Solution P2P Peer To Peer Performance Evaluation Quality of Service Simulation Study
Inspector
Console
Debugger
Style Editor
Performa...
Network
pper
div#ctl00_divLeftWrapper.left-wrapper
div.section-wrapper
div#divKeywordCloud.NormalTagCloud

ology of intrastriatal dopaminergic grafts determines functional and emotional outcome in n lesioned rats

Julia Jungnickel, Ieva Kalve, Linda Reimers, André Nobre, Maike Wesemann, Andreas Ratzka, Nina Halfer, Lindemann, Kerstin Schwabe, Kathrin Töllner, Manuela Gernert, Claudia Grothe

Journal: Behavioural Brain Research - BEHAV BRAIN RES, vol. 216, no. 1, pp. 129-135, 2011

A Practical Adaptive Pacing Scheme for TCP in Multihop Wireless Networks

Sherif M. ElRakabawy, Christoph Lindemann

Journal: IEEE/ACM Transactions on Networking - TON, vol. 19, no. 4, pp. 975-988, 2011

```
▸ <ul></ul>
  <div class="clear"></div>
▾ <div id="divKeywordCloud" class="NormalTagCloud">
  ▾ <span id="_tagCloudItemNew0" style="font-size: 12px; font-weight: normal">
      <a title="7 publication(s) related to this keyword" target="_self" href="/Keyword/1535/analytical-model">
      Analytical Model</a>
    </span>
  ▾ <span id="_tagCloudItemNew1" style="font-size: 15px; font-weight: bold">
      <a title="10 publication(s) related to this keyword" target="_self" href="/Keyword/60967/markov-chain">
      Markov Chain</a>
    </span>
  ▾ <span id="_tagCloudItemNew2" style="font-size: 14px; font-weight: normal">
      <a title="9 publication(s) related to this keyword" target="_self" href="/Keyword/25581/mobile-ad-hoc-
      network">Mobile Ad Hoc Network</a>
    </span>
  ▾ <span id="_tagCloudItemNew3" style="font-size: 12px; font-weight: normal">
      <a title="7 publication(s) related to this keyword" target="_self" href="/Keyword/28572/numerical-algorithm"
      >Numerical Algorithm</a>
    </span>
```

Figure: Example of the requirement of the HTML Meta Pattern

Candidate List Ranking

- Dou et al. [2] clusters similar lists together and ranks lists high if many result documents contain many terms of the list; they also require lists to appear and different websites
- Kong et al. [3] clusters terms of candidate lists into clusters based on their text and list context; afterwards he uses multiple TF and IDF measures to rank the facets
- Both do not penalize facets whose terms often appear together on each document
- Both do not differentiate between terms in lists and terms occurring on their own

Navigation Focused Idea

- Binary relevance assessment to decide if a specific facet term t is relevant for a specific search result document d: t is relevant for d if d contains t outside of lists $\rightarrow$ in this case t is a valid value for d in each facet that contains t
- Each facet term t induces a subset of the search results D'_t where t is relevant
- Idea: Measure the quality of the partition properties of the set of subsets $\{D'_{t1}, D'_{t2}, \ldots, D'_{tn}\}$ of facet $F = \{t_1, t_2, \ldots, t_n\}$ $\rightarrow$ facet extraction algorithm **NAV**

Search Result Pre-Processing

- Each search result document d is transformed into the bag of words representation $d' = \{t_1, t_2, \ldots, t_n\}$, containing only the terms not contained in lists → condensed document representation
- d' is generated at no cost: the candidate list extraction phase removes sub-tress / text snippets that contain the extracted list

$$D'_t = \{d' | d' \cap \{t\} \neq \emptyset\}$$

- We further define $D'_F = \bigcap_{t \in F} D'_t$ as the condensed search result

Facet Ranking Function

$$R_F = \alpha C_F + \beta S_F + \gamma P_F + \delta T_F$$

Partition Features I: Subtopic Coverage

- Subtopic coverage C_F recognizes the fact that the original query might have numerous interpretations, but each facet is only relevant for one of these possible search intents
- We approximate the number of sub-intents #I and calculate a distance measure to the expected number of documents matching at least one of the facet terms of F

$$\#I(D) = log(|D|)$$

$$C_F = \exp\left(-\frac{\left|\frac{|D|}{\#I(D)} - |D'_F|\right|}{10}\right)$$

Partition Features II: Size Equality

- S_F is a measure of the equality of the D'_t document set sizes with μ_F^S being the mean set size

$$\mu_F^S = \frac{\sum_{t \in F} |D'_t|}{|F|}$$

$$S_F = 1 - \frac{\sum_{t \in F} (\mu_F^S - |D'_t|)^2}{\sum_{t \in F} |D'_t|^2}$$

Partition Features III: Mean Number Facets

- The reciprocal of the mean number of facet terms per page P_F is used to prefer facets whose facet terms' co-occurrence rate is very low

$$\mu_F^C = \frac{\sum_{d' \in D'_F} |d' \cap F|}{|D'_F|}$$

$$P_F = \frac{1}{\mu_F^C}$$

Partition Features IV: Facet Size

- T_F is used to prioritize larger facets

$$T_F = \log |F|$$

Feedback Theory

- The feedback model defines how user selected facet terms are used to improve the web search result in terms of matching the user intent
- t^u represents a user-selected terms (feedback terms)
- $F^u = \{t_1^u, t_2^u, .., t_o^u\}$ is the set of feedback terms of facet F (feedback facet)
- $\mathcal{F}^u = \{F_1^u, F_2^u, ..., F_p^u\}$ is the set of non-empty feedback facets

Feedback Model

- Kong et al. [4] found Boolean filtering not useful
- They proposed soft ranking → original document score is combined with a score that depends on the feedback terms

$$S'_E(d, q, \mathcal{F}^u) = \lambda S(d, q) + (1 - \lambda) S_E(d, \mathcal{F}^u)$$

- Two implementations of S_E

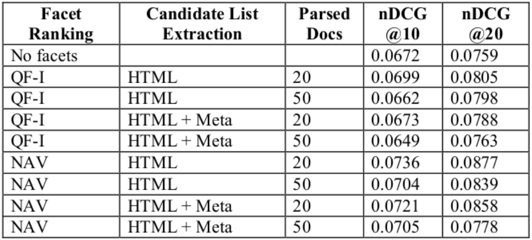
$$S_{ST}(d, \mathcal{F}^u) = \frac{1}{N} \sum_{F^u \in \mathcal{F}^u} \sum_{t^u \in F^u} S(d, t^u)$$

$$S_{TT}(d, \mathcal{F}^u) = \sum_{F^u \in \mathcal{F}^u} \sum_{t^u \in F^u} S(d, t^u)$$

Evaluation Model

- Extrinsic evaluation → impact an the search quality (NDCG)
- ClueWeb09 Category B corpus and TREC 2011 relevance measurements of the diversity task
  - Queries with sub-intents → relevance judgments for the sub-intents
  - Macro-Averaging: first average over the sub-intents per query, than over the queries
- Per sub-intent, incrementally add the remaining best facet term to the feedback terms
- BM25F ranking function

Single-Term Feedback I: Top-1 Facet

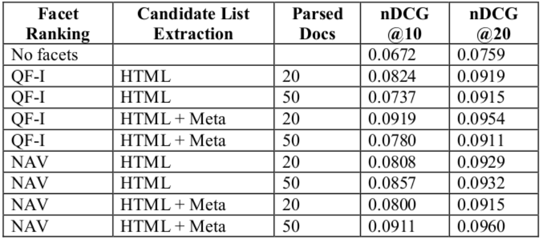


Facet Ranking	Candidate List Extraction	Parsed Docs	nDCG @10	nDCG @20
No facets			0.0672	0.0759
QF-I	HTML	20	0.0699	0.0805
QF-I	HTML	50	0.0662	0.0798
QF-I	HTML + Meta	20	0.0673	0.0788
QF-I	HTML + Meta	50	0.0649	0.0763
NAV	HTML	20	0.0736	0.0877
NAV	HTML	50	0.0704	0.0839
NAV	HTML + Meta	20	0.0721	0.0858
NAV	HTML + Meta	50	0.0705	0.0778

Figure: Single term feedback performance using top-1 facet

- NAV considerable higher scores than QF-I
- Meta Pattern impairs search quality

Single-Term Feedback II: Top-3 Facets

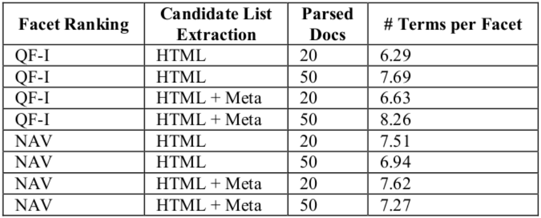


Facet Ranking	Candidate List Extraction	Parsed Docs	nDCG @10	nDCG @20
No facets			0.0672	0.0759
QF-I	HTML	20	0.0824	0.0919
QF-I	HTML	50	0.0737	0.0915
QF-I	HTML + Meta	20	0.0919	0.0954
QF-I	HTML + Meta	50	0.0780	0.0911
NAV	HTML	20	0.0808	0.0929
NAV	HTML	50	0.0857	0.0932
NAV	HTML + Meta	20	0.0800	0.0915
NAV	HTML + Meta	50	0.0911	0.0960

Figure: Single term feedback performance using top-3 facets

- NAV and QF-I achieve similar quality
- NAV requires top-50 documents to be effective
- Meta Pattern is beneficial

Single-Term Feedback III: Mean Number Facet Terms



Facet Ranking	Candidate List Extraction	Parsed Docs	# Terms per Facet
QF-I	HTML	20	6.29
QF-I	HTML	50	7.69
QF-I	HTML + Meta	20	6.63
QF-I	HTML + Meta	50	8.26
NAV	HTML	20	7.51
NAV	HTML	50	6.94
NAV	HTML + Meta	20	7.62
NAV	HTML + Meta	50	7.27

Figure: Mean number of facet terms of the top-3 facets

- Increasing number of documents is required to assess the NAV features correctly

Multi Term Feedback

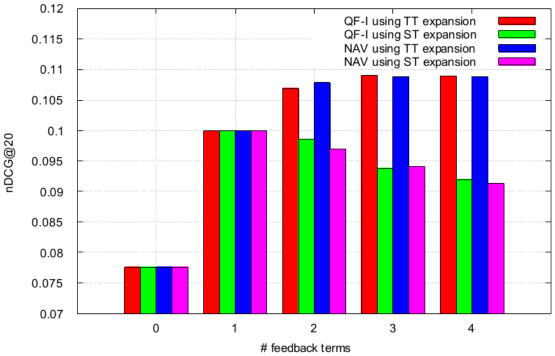


Figure: Results for multi term feedback using top-5 facets

- ST is not capable of utilizing more than one feedback term

Conclusion

- Facets generated by NAV, compared to QF-I facets, provide at least the same extrinsic utility
- Each baseline retrieval model might require its specific soft ranking expansion model
- Meta pattern HTML extraction algorithm yields lists that improve facet extraction significantly

Bibliography

Wisam Dakka and Panagiotis G Ipeirotis.
Automatic extraction of useful facet hierarchies from text databases.
In *Proc. ICDE*. IEEE, 2008.

Zhicheng Dou and et al.
Finding dimensions for queries.
In *Proc. CIKM*. ACM, 2011.

Weize Kong and James Allan.
Extracting query facets from search results.
In *Proc. SIGIR*. ACM, 2013.

Weize Kong and James Allan.
Extending faceted search to the general web.
In *Proc. CIKM*. ACM, 2014.

Chengkai Li and et al.
Facetedpedia: dynamic generation of query-dependent faceted interfaces for wikipedia.
In *Proc. WWW*. ACM, 2010.