"Elvis has left the building" Retrieval and Filtering of Events in Social Networking

Christos Tryfonopoulos

Assistant Professor Dept. of Computer Science & Technology University of Peloponnese

Outline



- Social Networks
 - challenges & motivation
- Retrieval of Events
 - social/semantic friendship
 - protocols & evaluation
- Filtrering of Events
 - user subscriptions
 - fast filtering & evaluation
- Past & ongoing work



The world has changed!







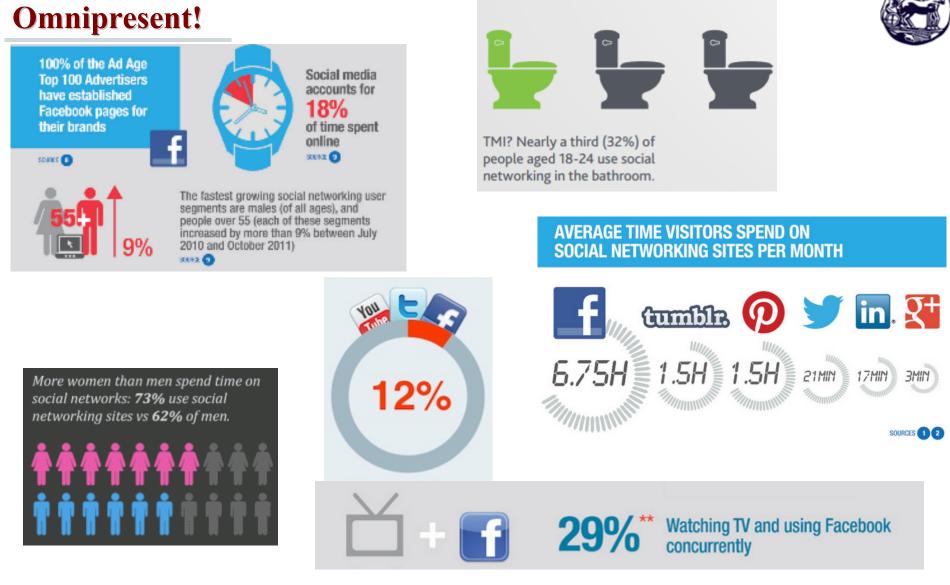


Retrieval and Filtering of Events in Social Networking – CS.UCy









AGES 18-24



Retrieval and Filtering of Events in Social Networking - CS.UCy

Social Networks



- The world is full of events!
 - "Elvis has left the building"
 - tweet/facebook status update
 - catch phrase for filtering tasks
- And we document almost everything!
 - and share it, comment on it, tag it
 - every single second...
- Prof. Mor Namaan put it correctly:
 - informers (25%)

VS

meformers (75%)





Μου αρέσει! · Σχολιάστε · Αποθηκεύστε την τοποθεσία · Κοινοποιήστε

Retrieval and Filtering of Events in Social Networking – CS.UCy

Social networks are fine, but...



- ... ever read the Terms of Service agreement?
- 13 pages, single-spaced 12pt font in Word
- 10 links to more specialised terms:
 - payments
 - marketing
 - reporting a violation
- Let me recap that for you:

"All Your Stuff is Ours, Even if You Quit"



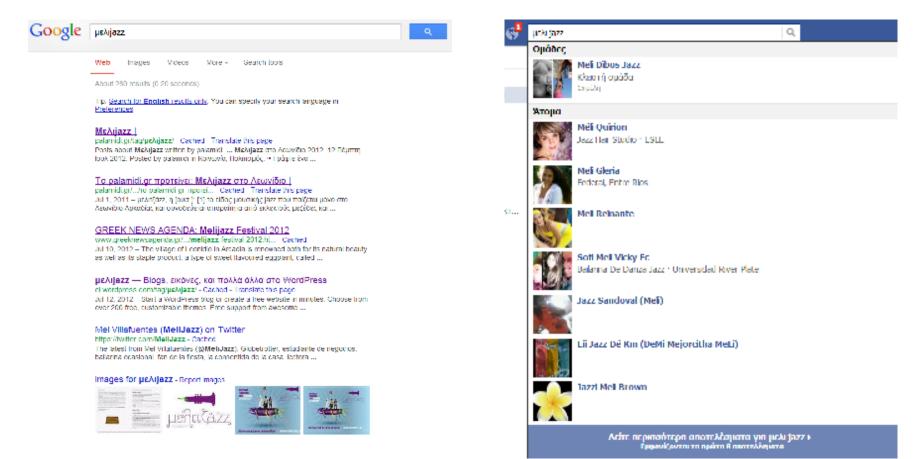
"Sign here to indicate you have no idea what you've signed."



Social networks are fine, but...



... ever tried to find something in Facebook?





Retrieval and Filtering of Events in Social Networking – CS.UCy

Social networks are fine, but...



- ... ever got bored to scroll down a timeline?
- "Δεσμοί" Κέντρο Έρευνας και Πρακτικών Εφαρμογών Αρχαίου Ελληνικού Δράματος
- Ακροατή μου: Για την Μπγιόρκ, τη Βενεζουέλα, τη γερμανική αποτελεσματικότητα και τον Μακουμπάνε
- Wish You Were Here (Full Album Vinyl Rip) Pink Floyd
- 03/04/2013 new member in family , baby girl at 13. 20
- Βαρκούλες αρμενίζουνε | Καθημερινές ιστορίες: χαμογελάστε παρακαλώ!
- ΑΤΜ=Άδικα Τα Μάζευα !!!



Motivation



- ... ever read the Terms of Service agreement?
 - your content, should be yours! ⇒ "Give the data back to the people"
 - break information monopolies
 - self-* properties (self-manageable, self-healing, self-organizing, ...)
- ... ever tried to find something in Facebook?
 - make events/content/data available
 - exploit small-world phenomenon
 - there are other friendship types: e.g., users sharing common interests!
- ... ever got bored to scroll down a timeline?
 - information overload ⇒ missed events
 - subscribe to interesting events (alerting/filtering!)



Outline

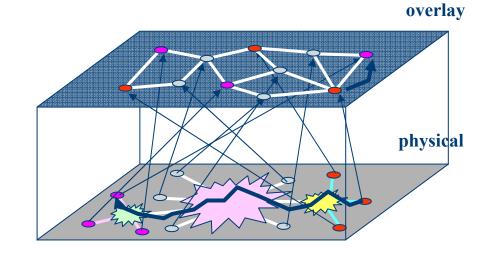


- Social Networks
 - challenges & motivation
- Retrieval of Events
 - social/semantic friendship
 - protocols & evaluation
- Filtrering of Events
 - user subscriptions
 - fast filtering & evaluation
- Past & ongoing work



Decentralised Social Networks

- An overlay network is built
- Each user node stores links to social friends (explicit)
- Connected nodes are called friends (one hop away)
- Nodes use their links to perform social tasks
 - poke,
 - status updates,
 - •







So far...



- Distributed Social Platforms
 - Diaspora, KrawlerX, OpenSocial, SuperNova, Scope
 - servers managed by single stakeholder
 - two-tier architecture, higher-level nodes belong to a single provider
 - LifeSocial.KOM, SCIMS, GEMstone
 - emphasise on pokes/ontological descriptions of friendship, not on content
- Decentralised Social Networks
 - My3, P³D
 - emphasis on security/privacy of interactions
 - eXO, SocialCDN
 - emphasise message traffic, impose strict structure (e.g., DHT)



Social links only are not enough!



- Our friends are OK for
 - pokes
 - status updates
 - vacation photos
- But they suck at content retrieval tasks!
 - try to find a piece of information/event/document in Facebook!
 - we need more elaborate retrieval mechanisms
- We believe that our friends have what we may need
 - NOT true!
 - I'll show you that they have only 30% of what we need ;)



Node Structure



DS⁴: A Distributed Social and Semantic Search System

Each node in the social network has 2 routing tables:

- Friends Index (FI) straightforward
 - good for the basic social networking services
 - used for
 - explicit social friends
 - reaching offline content through FOAF links
- Semantic Index (SI) more interesting!
 - good for sophisticated content/event/object search & retrieval
 - used for
 - semantic friends (implicit), aka friends with similar interests
 - semantic tasks (content retrieval)

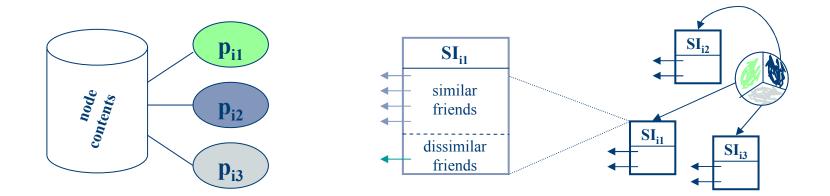




Node Join



- When a node n_i joins the network for the first time
 - 1. (semi-)automatically categorises its content
 - 2. computes its interests/node description called persona p_i
 - 3. for each persona p_{ik} it maintains a Semantic Index SI_{ik} SI_{ik} is initialised by friends (FI links), but this will change!
- These (randomly selected) links will be refined

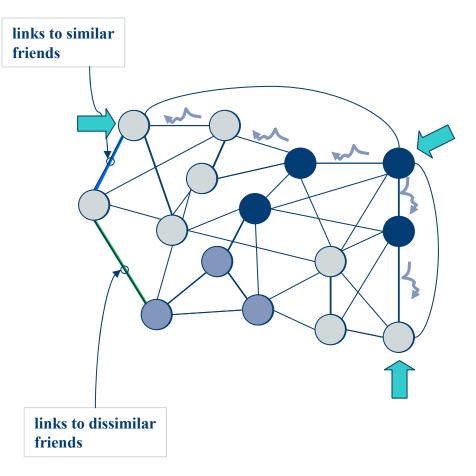




Discovering Semantic Friends



- Viral, fully-autonomous process
 - initiated independently by each node
 - based on local view/friends
 - periodic heart-beat protocol
 - abandon old/outdated/dissimilar friends for
 - fresh/up-to-date/similar ones
- Self-organising communities of friends
 - create small-worlds of semantically close nodes
 - all interesting information a few hops away!

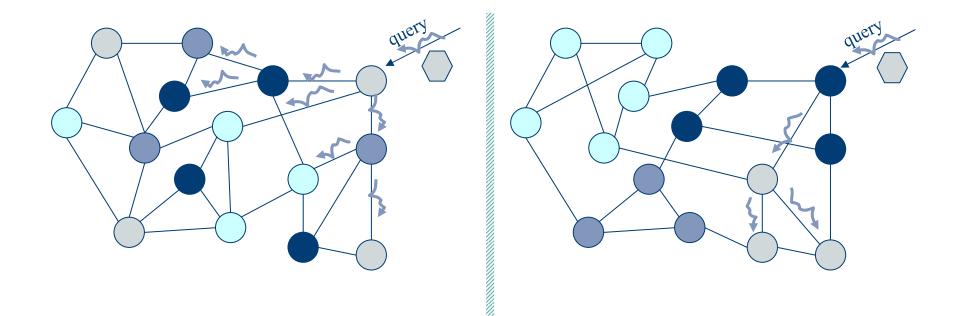




Content retrieval

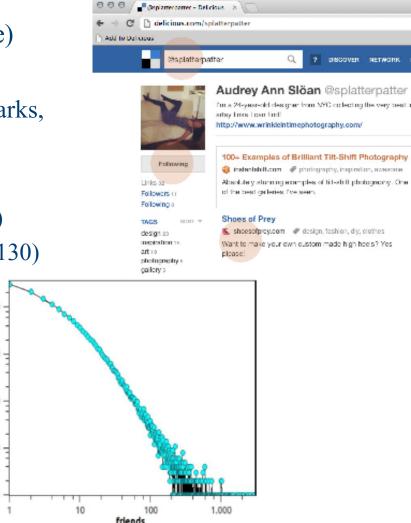


- Locate the group/cluster of nodes similar to the query
 - once they reach such group, they are broadcasted ("fireworks")
- Exploit the organisation of the network
 - more and better results, less time, less hops, less messages!

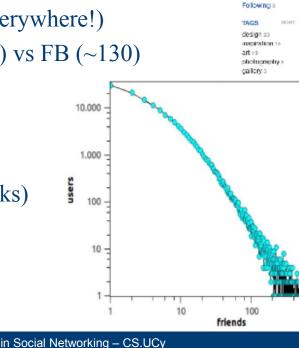


Experiments

- Delicious (social bookmarking service) home crawl
 - 700K users, 10M tags, 21M bookmarks, 170K topics/categories
 - graph:
 - power law (actually everywhere!)
 - small # of friends (~11) vs FB (~130)
 - overlap:
 - 15% in friends
 - 17% in tags
 - 0,6% in data (bookmarks)
 - 0,4% in categories

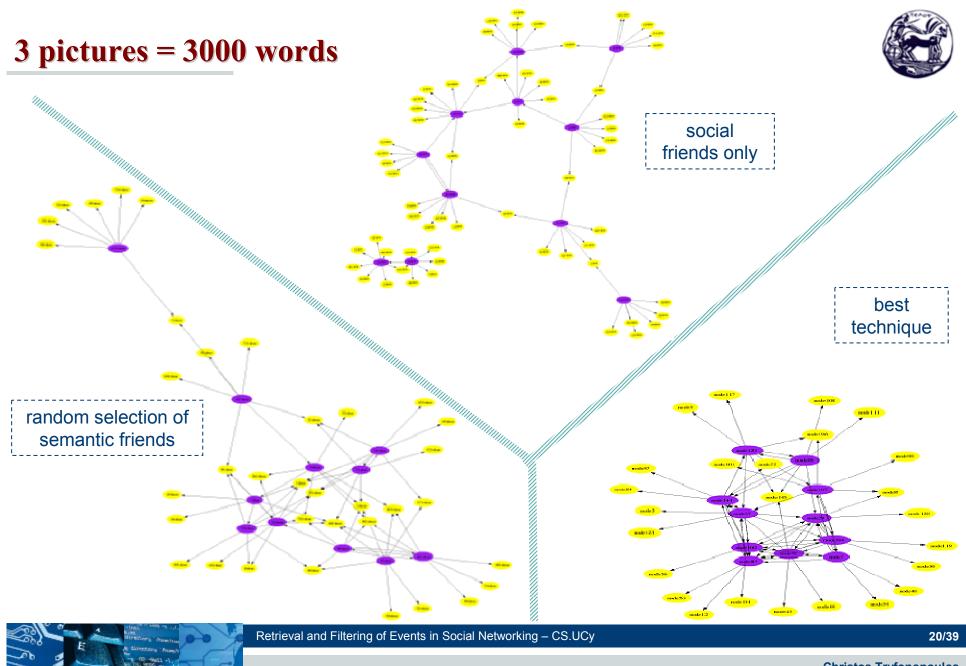








Retrieval and Filtering of Events in Social Networking - CS.UCy



Recall

MSP _____

RF

MSF

Experiments

0.9

0.8

0.7

0.6

0.5

0.4

0.3

0.2

0.1

0

recall



• For fans of graphs :)

Message traffic/query 140 120 100 80 60 MSP MSFP MSF 40 RF 20 0 10 20 30 40 50

time units



Retrieval and Filtering of Events in Social Networking – CS.UCy

50

40

searcn messages (x 10)

Outline



- Social Networks
 - challenges & motivation
- Retrieval of Events
 - social/semantic friendship
 - protocols & evaluation
- Filtrering of Events
 - user subscriptions
 - fast filtering & evaluation
- Past & ongoing work



The other side of the coin



- Searching or one-time querying or (simply) querying
 - a user poses a query "I want information for the Meli-Jazz festival".
 - the system returns matching content: tweets, webpages, photos
- Alerting or publish/subscribe (pub/sub) or filtering or persistent querying or continuous querying or information push
 - a user posts a continuous query or profile or subscription
 - created explicitly from the user or
 - implicitly (e.g., a user-profiling software agent)
 - the new Meli-Jazz festival is announced!
 - the system notifies the user with matching content: tweets, webpages, photos



Filtering in a nutshell



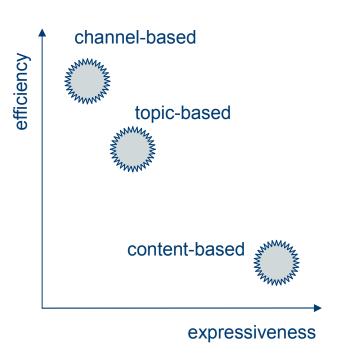
- First generation filtering systems
 - channel-based (aka group-based)
 - a set of groups designated by the system
 - each event published to one such group
 - user subscribes to one or more groups of interest
 - \rightarrow think of mailing lists or IP multicast
 - topic-based (aka subject-based)
 - (a bit) more flexible
 - each event is tagged with a subject (from a vocabulary or arbitrary)
 - user subscribes by specifying the subject (and operations $*,?,\vee,\wedge,...$)
- fast filtering, simple implementation but ...
 - no flexibility, cognitive overload to user



Filtering in a nutshell



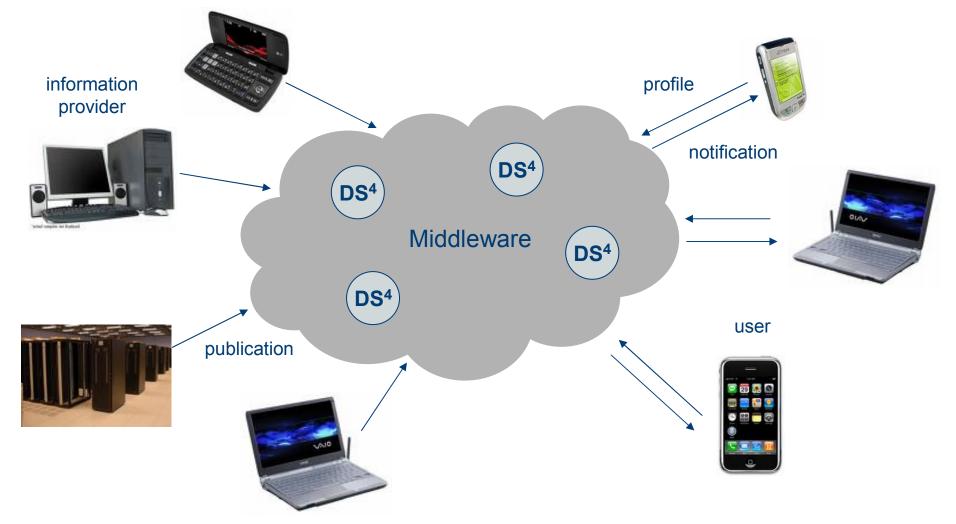
- Second generation filtering systems
 - content-based (our focus)
 - rich data models
 - users specify arbitrary profiles
 - matching relies on the subscriber, not on the publisher
- but ...
 - more complex
 - specialised data structures, algorithms
 - more expensive
 - time, computational effort





A Filtering Scenario: Social Networking







Retrieval and Filtering of Events in Social Networking - CS.UCy

Filtering in social networks



- According to Facebook, more than 50% of the users are through mobile devices (monthly: 1Bn total, 543M mobile)
- Filtering is also suitable for mobile environments
 - asynchronous (does not require an uninterruptible connection)
 - multicasting in nature (can reach many devices)
 - dynamic (copes with often connections/disconnections)
- Filtering: identifies a new type of user, the subscriber:
 - non-immediate information or service need
 - events of interest might include: new iphone release, 10 of my friends liked an article, shirts under 50 euro around the corner...
 - strong connection with location, timeliness, availability, device capabilities, ...



User subscriptions



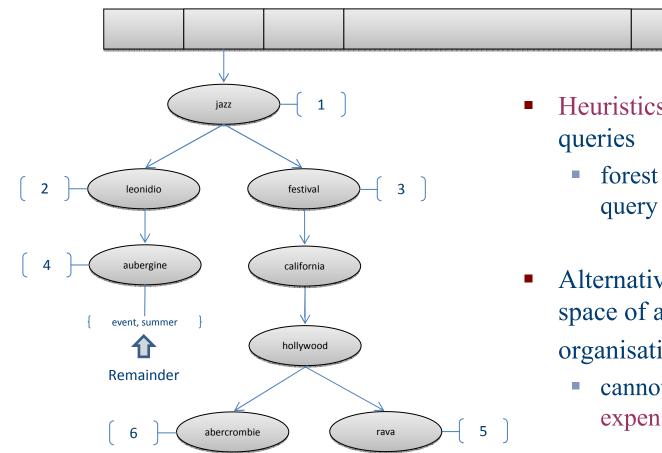
- I want to get notified when an event/document is published that
 - has been tagged with the tags jazz, festival, leonidio and/or
 - contains the words jazz, event, rava

- How to index queries?
 - no index, examine them all \rightarrow come on...!
 - inverted index \rightarrow too slow
 - cluster them!



Query indexing





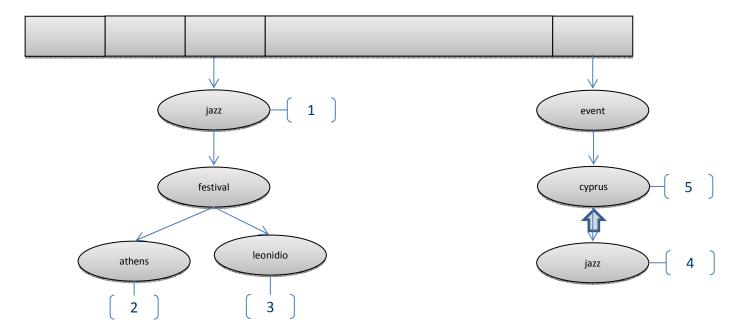
- Heuristics to organise the queries
 - forest of tries depends on query insertion order
- Alternative: search over the space of all possible query organisations!
 - cannot be done, way too expensive!











Can we do a better job in clustering?

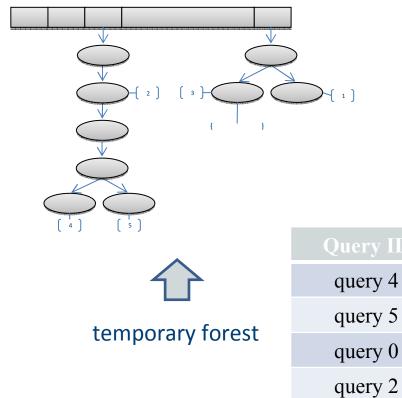


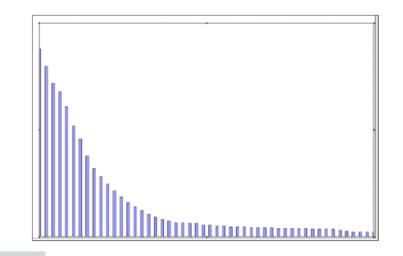
Retrieval and Filtering of Events in Social Networking – CS.UCy

Can we do even better?



YES! Use statistics of tags





Determine a new insertion order (maximise benefit)

Retrieval and Filtering of Events in Social Networking – CS.UCy

query 3

query 1

14

14

11

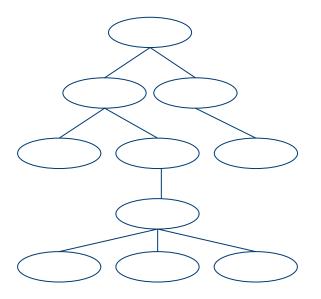
9

8

6

How to filter each new event/document/tweet/post?

- Examine as few queries as possible
 - search in selected tries only
 - prune tries when searching
 - clustering matters!
- How fast is this?
 - 6 times faster than inverted index...



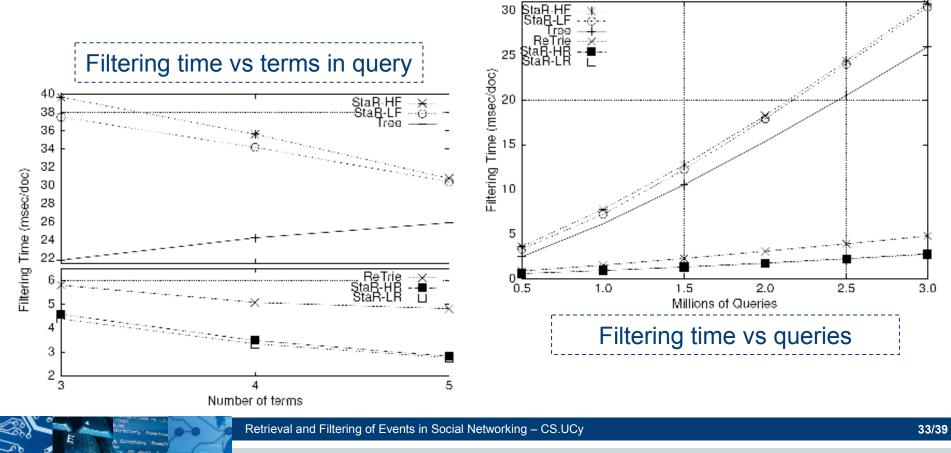




Experiments

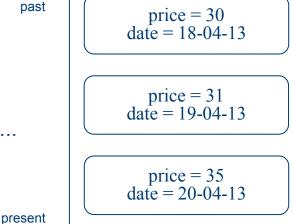


- Delicious tags and bookmarks as docs
- DBpedia long abstracts (3M documents, 2.8GB, >800K unique terms)



Other (important) issues

- Mobile devices disconnect frequently (battery, communication problems, lack of signal reception), users logout
- What happens to user notifications?
 - stored to deliver upon reconnection
 - how many, for how long?
 - which ones?
 - some may be subsumed by others, outdated, ...
- When to deliver notifications?
 - might not make sense any more
 - timing constraints
 - batching vs forwarding





34/39

Conclusions



• Next generation search \rightarrow filtering ("identify the need without a query")

In social/mobile environments:

- subscribe with a query
 - implicitly (through a software agent)
 - explicitly (by the user)
- refine and augment queries

In advertising:

- filtering = opportunity awareness
- active, targeted advertising (ads, not passive any more)



Outline



- Social Networks
 - challenges & motivation
- Retrieval of Events
 - social/semantic friendship
 - protocols & evaluation
- Filtrering of Events
 - user subscriptions
 - fast filtering & evaluation
- Past & ongoing work



Work you may find interesting



- Semantic/Personal Information Management
 - [ISWC'11, ESWC'12]
- Distributed/P2P Information Retrieval/Filtering
 - [ECIR'13,Coopis'11, DAPD'09, Intern. Comp.'07, WISE'08, SIGIR'05 & '08, P2P'08]
- Digital Libraries
 - [JCDL'09, ECDL'08,'07,'05, DELOS'07]
- Distributed DBs
 - [ICDE'06, SIGMOD'04, EDBT'04]
- Security & privacy in P2P
 - [PODC'08, CoopIS'09]
- Data models/query languages/local indexing algorithms
 - [TOIS'09,TKDE'06, SIGIR'04, SIGMOD Record'03, ECDL'02]



Ongoing work



- Decentralised online social networks
 - architectures and cross-platform service provisioning
 - novel services
 - distributed social content search, aggregation functions, filtering
 - privacy-protecting APIs
- Semantic Web
 - filtering streams of RDF data
 - subscribing to ontology changes
- Personal Information Management
 - semantic browsing & searching for the desktop
 - mindmaps for content visualisation



Thank you...

... for your attention

Also thanks go (among many others) to:

- Manolis Koubarakis
- Gerhard Weikum
- Stratos Idreos
- Christian Zimmer
- Paraskevi Raftopoulou
- Yannis Drougas
- Matteo Maffei
- Alessandro Linari
- Euripides Petrakis
- Lefteris Zervakis

- @ University of Athens
- (a) Max-Planck Institute
- @ CWI
- @ Max-Planck Institute (Future Consulting)
- @ University of Peloponnese
- @ ESRI
- @ MMCI & University of Saarland
- @ Nominet
- @ Technical University of Crete
- @ University of Peloponnese