Comparing PubSubHubbub and Twitter

Introduction to Information Retrieval
CS 221
Donald J. Patterson

PubSubHubbub

- Basic idea is to eliminate polling sites for changes
  - = eliminate crawling web sites blindly

PubSubHubbub

- A feed declares its PuSH server in it’s header
  - A feed is a URL
  - A feed is an Atom or RSS XML file
  - A PuSH server is a “hub”

  -
    ```
    <link rel="hub" href="http://myhub.example.com/endpoint" />
    ```
  - A PuSH server is a “hub”

PubSubHubbub

- A subscriber first gets the RSS feed as always

PubSubHubbub

- A subscriber then subscribes to the “hub”
  - The hub then tells the subscriber when there is an update
  - This avoids endless polling by the subscriber for changes

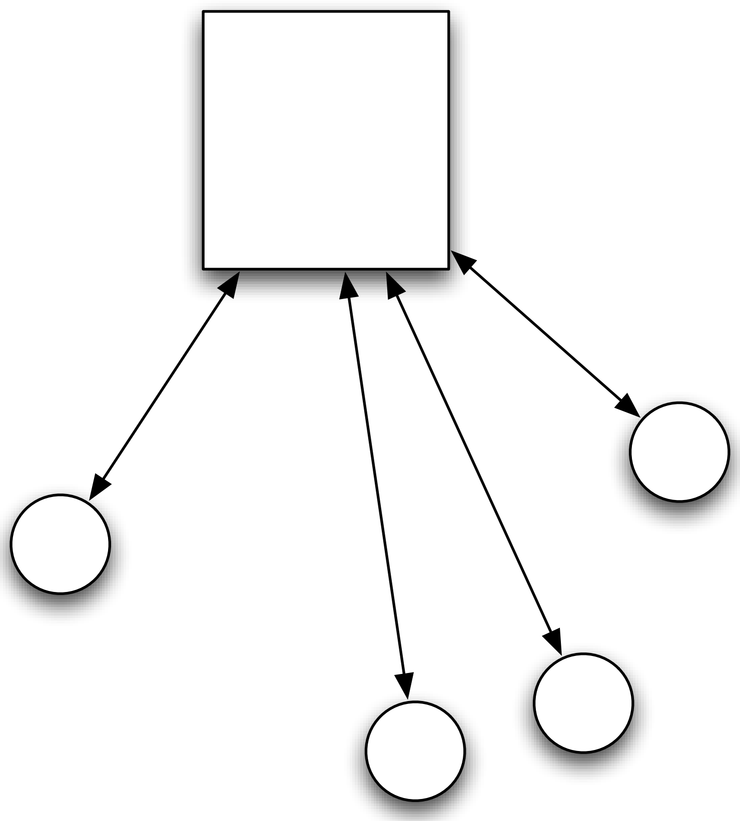
PubSubHubbub

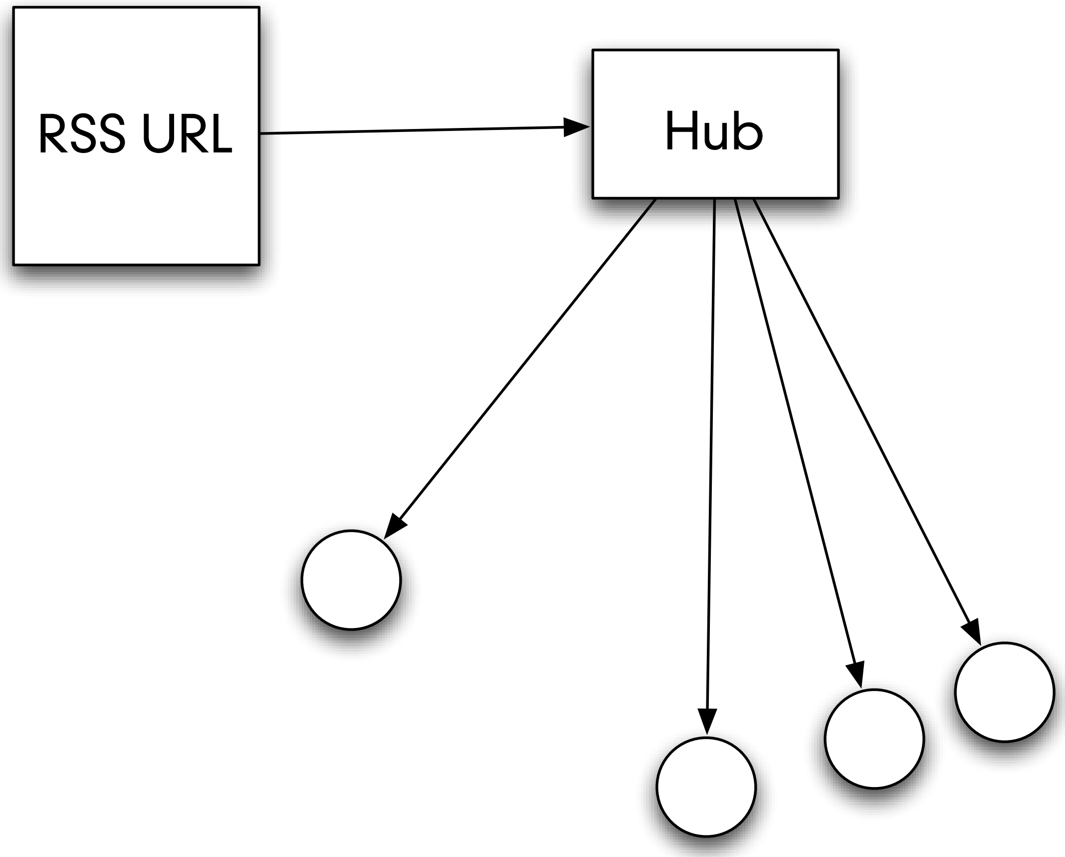
- When there is an update, the publisher tells the hub

PubSubHubbub

- The hub gets the feed and sends it to the subscribers

PubSubHubbub





PubSubHubbub

http://docs.google.com/present/view?id=ajd8t6gk4mh2_34dvbpchfs

http://www.youtube.com/watch?v=B5kHx0rGkec

Twitter API

- In contrast to PuSH
  - Twitter takes control of both publish and subscribe
  - Twitter becomes a black box for multicast
  - PuSH decouples publication from subscription
    - PuSH is much more scalable
    - Not as popular

Twitter API

- There are actually three APIs
  - REST interaction
  - REST search
  - Stream-based
    - The Streaming API provides low-latency high-volume access to Tweets.

Twitter API

- You cannot make unlimited calls, follow requests, updates or direct message
  - API usage is rate limited.
  - There are limits on the number of follow requests, updates and direct messages you can make in a single day.

Twitter API

- The API is entirely HTTP-based
  - Methods to retrieve data from the Twitter API require a GET request.
  - Methods that submit, change, or destroy data require a POST.
  - API Methods that require a particular HTTP method will return an error if you do not make your request with the correct method

Twitter API

- The API is a RESTful resource
  - Simply change the format extension a request to get results in the format of your choice.
  - The API presently supports the following data formats: XML, JSON, and the RSS and Atom syndication formats, with some methods only accepting a subset of these formats.

Twitter API

- Parameters have certain expectations
  - Some API methods take optional or requisite parameters.
  - Parameter values should be converted to UTF-8 and URL encoded.
  - The page parameter begins at 1, not 0.

Twitter API

- There are many libraries available

Twitter API

- Examples:
  - Get the public timeline in RSS format, unauthenticated

    -
      ```
      curl http://api.twitter.com/1/statuses/public_timeline.rss
      ```
  - Get updates from users you follow in XML, authenticated

    -
      ```
      curl -u username:password http://api.twitter.com/1/statuses/friends_timeline.xml
      ```
  - Post a status update and get the resulting status back as JSON

    -
      ```
      curl -u username:password -d status="your message here" http://api.twitter.com/1/statuses/update.json
      ```

Twitter API

- Examples:
  - Streaming API

  -
    ```
    curl -d @locations http://stream.twitter.com/1/statuses/filter.json -u username:password -s
    ```

Discussion

- Discussion