Getting Started With Condor
Contents

- Getting Started
- Collecting Web Content
- OneDegreeCollector
- Building your own Startlists
- Collecting your E-Mail
- Collecting Facebook Data
- Collecting Wikipedia Data
- Collecting CoolPeople
- Coolhunting Blueprint
Getting Data into Condor

- IMAP E-Mail (Mailcollector) - Eudora mailboxes
- Web/Blog/News/Scholar (WebCollector)
- Wikipedia (Wiki-Search, WikiFactFetcher)
- Snippets (OneDegreeCollector)
- Twitter (TwitterCollector)
- FlatFiles (FileLoader)
- PeopleNetworks (CoolPeople, MyNet)
- Facebook

Communication View (social net)
Term view (semantic net)
Communication View (link net)
Term view (semantic net)
Communication View (semantic net)
Term view (semantic net)
Communication View (social net)
Term view (semantic net)
Communication View (social net)
Term view (semantic net)
Communication View (social net)
Temporal Visualization by a Sliding Time Frame

With history:

- Time Frame size $n$

no history:

- Time Frame size $n$
With and without history
Preparation

- Install MySQL
- Install Java (only Windows)
- Install Java 3D (only Windows)
- Start Java (if it does not run yet)
Starting Condor

- java -Xmx2048M -jar condor-2.1.4.jar

Condor Key
MySQL password
(default: no password)
Tools to collect data

- Web Collector
- One Degree Collector
- Mail Collector
- WikiFactFetcher
- Twitter Collector
- Facebook Fetcher
- Video Collector
- File Loader
- Cool People
- My-Net
- Wiki-Search
Contents

- Getting Started
- Collecting Web Content
- OneDegreeCollector
- Building your own Startlists
- Collecting your E-Mail
- Collecting Facebook Data
- Collecting Wikipedia Data
- Collecting CoolPeople
- Coolhunting Blueprint
Getting a Bing API Key

Develop with Bing
Use Bing Search and Map APIs to increase your website's function and applications.

Sign in - Bing Search API
Sign up to use Bing Search API and create AppIDs
Learn more

Create unique search applications
- Build applications powered by Bing's technology
- Choose from multiple SourceTypes (Web, Images, Video, and More) and output protocols (JSON, SOAP, or XML)
- Customize the search results to your needs

Create your Bing Webmaster
RESOURCES
- Blog
- Forums
- Documentation
- Terms of use

APPLICATION
Bing API offers open, flexible options for building or enhancing your site or applications. Developing an application with the new API is straightforward. Choose a SourceType or SourceTypes you're not limited to one), choose an output protocol (JSON, SOAP, or XML), and then customize according to your needs.

HOW TO GET STARTED
- Create your AppID
- Download Bing API SDK for Search or Maps
- Build your solution with Bing Maps
Collect Web Content
Communication View

Create Communication Views

- View all selected datasets
- Record view loading process into a file
- Set time constraints on communication data for the view
- Refine site
- Refine data by Actor’s Communication Frequency

Static View of Communications

- Dynamic View of Communications
- Netgraph view of Communications
- Local Context View of Communications
- Prepare Gossip View

[Diagram of communication network with nodes labeled: hillary clinton, mediamatters.org, john mccain, john edwards, mitt romney]
Term view index

- Use existing index (uncheck to create a new index)

- Terms Selection Strategy
  - Preview terms
  - Upload list of terms

- Indexing
  - Include numbers (whether numbers should be indexed)
  - Use Porter stemming (whether words should be reduced to their stem)
  - Set constraints on document date

- Stopwords
  Stopwords are highly frequent words like "the" or "for". Stopwords are skipped during...
  - use minimal stopwords list
  - use default stopwords list
  - use extended stopwords list
  - use blog stopwords list
  - use custom stopwords list

  (one word or phrase per line, no commas, lowercase words)

Continue
Term view index - 2

This dialog lets you select which terms will be used for the term view.
Click the words or phrases from the left column to add it to the right selection.

Add Words
word (norm. frequency) (click to add)
- v6.00.2800.1165 (56.0)
- mimeole (56.0)
- priority (56.0)
- microsoft (56.0)
- msmai (56.0)
- charset (32.0)
- iso-8859-1 (32.0)
- http (29.0)

Add 10 from the top

Add Phrases (norm. frequency) (click to add)
- microsoft before (56.0)
- microsoft very (56.0)
- msmai before (56.0)
- msmai very (56.0)
- charset very (32.0)
- charset before (32.0)
- http before (29.0)
- http very (29.0)

Add 10 from the top

Selected Terms
(click item to remove)
- produced (56.0)
- survey (36.0)
- thomas (31.0)
- mobile (30.0)
- deloitte (28.0)
- rico (28.0)
- ckn (28.0)
- consulting (28.0)
- sent (26.0)
- schmalberger (26.0)
- web (26.0)
- please (26.0)
- sent before (26.0)
- sent very (26.0)
- web before (26.0)
- web very (26.0)
- before mailing (26.0)
- very mailing (26.0)
- 2002 (25.0)
- thurgauerstrasse (24.0)
- converted (24.0)
- pglm (24.0)
- original (24.0)
- zurich (24.0)
- users (24.0)
- boundary (24.0)
- google (22.0)

Sort terms by norm. frequency
- sort terms by name

- Factor actor's weight
  in when computing communication weight
- Index dates
  (necessary for dynamic termview)
- Keep top % of cococ. 0.5

All your previous terms
selections are automatically stored.
- Use previous selections
- Only Export selected terms
- OK, collect communications ...
Contents

- Getting Started
- Collecting Web Content
- **OneDegreeCollector**
- Building your own Startlists
- Collecting your E-Mail
- Collecting Facebook Data
- Collecting Wikipedia Data
- Collecting CoolPeople
- Coolhunting Blueprint
One-Degree-Collector

- Complementary to the Blog Collector
- Fetches only one degree
- Retrieved websites are not aggregate
One-Degree-Collector - UI

- GUI resembles Blog Collector
One-Degree-Collector - result

- typical result of one-degree search
Contents

- Getting Started
- Collecting Web Content
- OneDegreeCollector
- **Building your own Startlists**
- Collecting your E-Mail
- Collecting Facebook Data
- Collecting Wikipedia Data
- Collecting CoolPeople
- Coolhunting Blueprint
Creating Term View Without OneDegreeCollector Start

List: Create Stoplist First
... then use this stop list for the term view
Creating Term View With Start AND Stop List

- Use existing index (uncheck to create a new index)

- Preview terms

- Upload list of terms: elcat/ArtRelCat-2.08/bankingstoplist

- Include numbers (whether numbers should be included)
- Use Porter stemming (whether words should be reduced to their stems)
- Set constraints on document date

Stopwords
- Minimal stopwords list
- Default stopwords list
- Extended stopwords list
- Blog stopwords list
- Use custom stopwords list: elcat/ArtRelCat-2.08/bankingstoplist

Graph showing term view with various nodes and connections.
Contents

- Getting Started
- Collecting Web Content
- OneDegreeCollector
- Building your own Startlists
- Collecting your E-Mail
- Collecting Facebook Data
- Collecting Wikipedia Data
- Collecting CoolPeople
- Coolhunting Blueprint
Left side: enter here the specification of the mailbox

Right side: database related data, eg. username: root, no password

For username, host, port, and ssl check with your email provider (for gmail, see next slide)

Content: yes will download the whole emails, w/o content only the sender, recipients and the subject line are downloaded

Anonymize will replace email addresses with random identifiers

Here you can choose specific folders to download

Delete the present data in the database?
Settings for gmail

Don’t forget the access information for your mysql database on the right, then press start. It might take a while (esp. with huge mailboxes) before you see a progress bar.
Visualize Mail-Data

1. Open Database...
2. Database Name: information_schema
3. Connected to database petermailDec09@localhost. Communication type is: For Emails type
4. View all selected datasets
   - Record view loading process into a file
   - Set time constraints on communication data for the view
   - Refine mailAddress
   - Refine data by Actor's Communication Frequency
Visualize E-Mail Data (2)
Visualize E-Mail Data (3)

**Input**

keep actors whose communication frequency >= your input. The range is [1 - 9100]

- enter your input value (e.g., 2)
- cancel or OK

**Dynamic Control**

- Time Unit: DAY
- Time Window Size [1, 61]: 1
- With History
- Fast Decay
- Add temporal attributes (only w/out History)
- OK

**Select top N nodes by betweenness**

- total nr of nodes: 311
- group density of graph: 0.01550689
- Select top N nodes by BC
- top N nodes: 311 (min. 0, max. 311)
- Extrapolate removed nodes over 2 degrees
- Weight extrapolated edges with factor: 0.25
- Extrapolate removed nodes over 3 degrees
- Weight extrapolated edges with factor: 0.1

**galaxyadvisors**
Dynamic View of Communication
Visualize E-Mail Contents

1. View, Tools, Window, Help
2. Communication, Term, Documents
3. Content Process, Static View of Terms, Dynamic View of Terms, Taxonomy of Terms

Term processing

- Use existing index (uncheck to create a new index)
- Terms Selection Strategy
  - Preview terms
  - Upload list of terms
- Indexing
  - Include numbers (whether numbers should be indexed)
  - Use Porter stemming (whether words should be reduced to their stem)
  - Set constraints on document date
- Stopwords
  - Stopwords are highly frequent words like "the" or "for". Stopwords are skipped during...
    - use minimal stopwords list
    - use default stopwords list
    - use extended stopwords list
    - use blog stopwords list
    - use custom stopwords list
  - (one word or phrase per line, no commas, lowercase words)
Visualize E-Mail Contents (2)
Dynamic View of Terms
Contents

- Getting Started
- Collecting Web Content
- OneDegreeCollector
- Building your own Startlists
- Collecting your E-Mail
- Collecting Facebook Data
- Collecting Wikipedia Data
- Collecting CoolPeople
- Coolhunting Blueprint
Collecting Wikipedia Data
Output of Wikipedia