

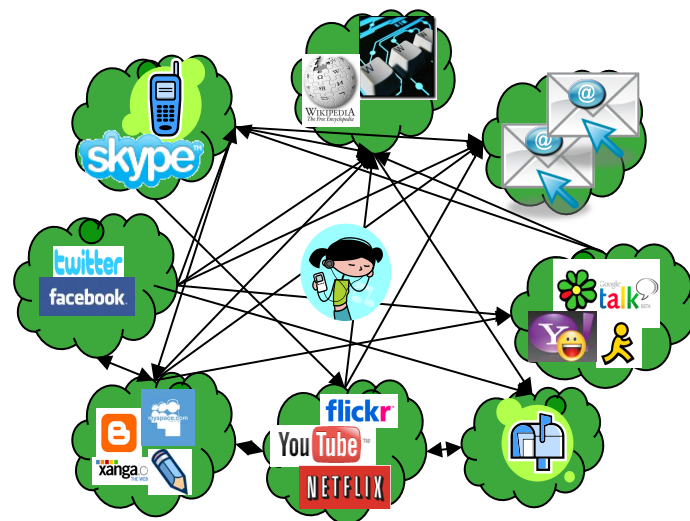
Hone

Automatically Watching Across Information Networks

Smriti Bhagat & S. Muthukrishnan
(CCICADA, Rutgers Univ.)
Corporate Partner: Narus Inc.

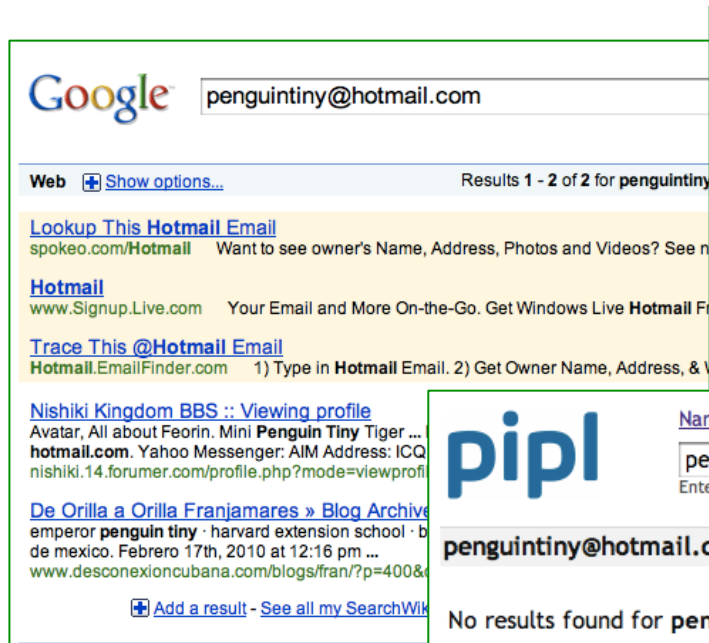
Problem

- Individuals have multiple online IDs
- **Given:** a set of IDs to be tracked (e.g., email ids)
- **Task:** find information about people behind IDs across networks (e.g., other email ids, mobile#, address, facebook friends, VoIP calls)

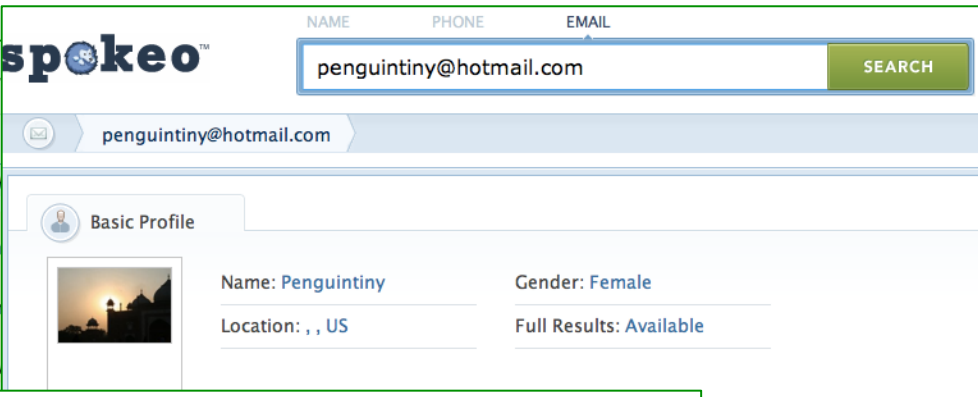


Example

Given: email id (penguintiny@hotmail.com)

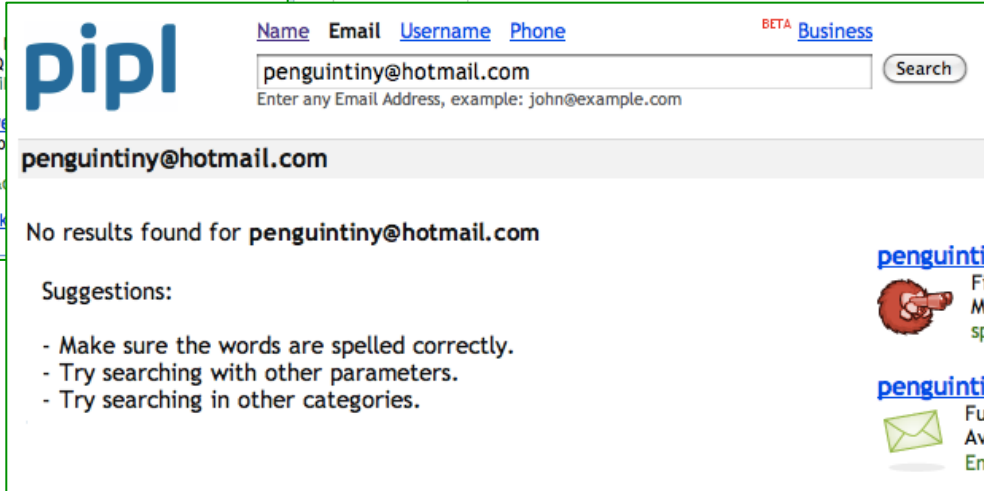


Google search results for **penguintiny@hotmail.com**. The results show several links related to the email address, including "Lookup This Hotmail Email" from Spokeo, "Hotmail" from www.Signup.Live.com, "Trace This @Hotmail Email" from Hotmail.EmailFinder.com, and "Nishiki Kingdom BBS :: Viewing profile" from nishiki.14.forumer.com. There are also links to "De Orilla a Orilla Franjamares » Blog Archive" and "Add a result - See all my SearchWiki".



Spokeo search results for **penguintiny@hotmail.com**. The search bar shows the email address and a "SEARCH" button. Below the search bar, the "Basic Profile" section displays the following information:

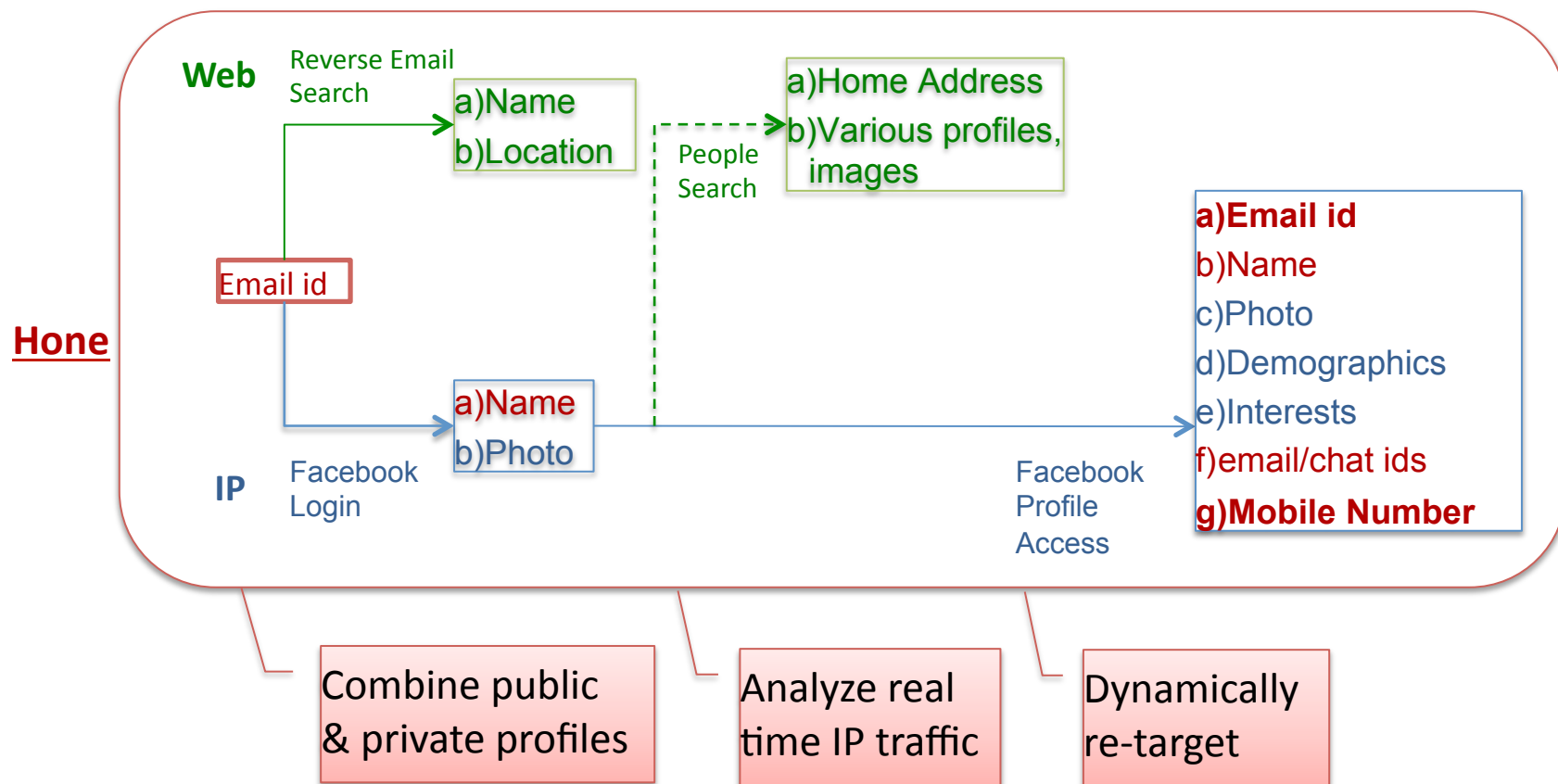
NAME	PHONE	EMAIL
Name: Penguintiny		
Location: , , US		
Gender: Female		
Full Results: Available		



Pipl search results for **penguintiny@hotmail.com**. The search bar shows the email address and a "Search" button. Below the search bar, the results section displays "No results found for penguintiny@hotmail.com". The "Suggestions" section lists the following:

- Make sure the words are spelled correctly.
- Try searching with other parameters.
- Try searching in other categories.

Demo: Example



Challenges

- **Discover:**
 - RegEx matching to **filter** relevant IP traffic at wire speed (Narus:40Gb/s)
Cookie:[]+[a-zA-Z0-9_-=?;|.|()]*c_user=[0-9]*[a-zA-Z0-9_-=?;|.|()]*lxe=email_id
 - Parse reconstructed IP session for private information (tens of Mb/s)
 - Selectively query/crawl public networks (tens of profiles/s)
- **Analyze:**
 - Build equivalence classes of ids
- **Hone:**
 - Feedback loop to proactively:
 - update watchlist with RegExs,
 - re-crawl public web and re-target, and
 - update equivalence classes

Ongoing work

- Algorithms
 - Faster RegEx filtering on IP Traffic
 - Probabilistic equivalence among ids
- Engineering
 - Javascript/flash based web parsing
- Extend functionality
 - Topics/communities to watch
 - Geo-location info (web, mobile)
- Find partners

Thank you!