Find and Study User Actions from HTTP Traces

**User Actions** = webpage URLs explicitly requested by a user

1. Find User Actions from HTTP traces
2. Analyze User Actions
3. Model user navigation behaviour

- Show how users surf the Web
- Build model of real users to:
  - Forecast their behaviors
  - Enhance webpage ranking algorithms

1 - Find User Actions

Users in households ➔ Network traffic passively observed ➔ Pageview URLs

PoP

Data from 20k users, 3 years

2 – Analyze User Actions

- Quantify visited webpages and domains
  - Paths followed by the users
  - Clusterization of the browsing behaviour
- Impact of devices
- Identification of prominent content promoters
- Role of social networks and search engines

Examples:

- Time between two User Actions
  - 2-3 4-5 6-7 8-9 10-11 12-13
  - Webpages weekly visited per device
  - Webpages and domains per path

Compare Hot and Cold Web

- Hot Web: part of Web visited through User Actions
- Cold Web: active crawling of Web static structure

Built an automatic multi-threaded crawler to obtain Cold Web

- Many webpages quickly disappear
- Most of the hyperlinks are never followed
- Hot Web is based on search engines and social networks, with personalized and dynamical pages

3 - Model user behavior

Random surfer model ➔ PageRank

Likelihood that a user randomly walking will arrive at a certain webpage

Popularity of webpages for random surfers

Comparing with User Actions, random surfer model is not adequate

---

Luca Vassio
luca.vassio@polito.it
luca.vassio.wordpress.com

Supervisor: Marco Mellia