



Relevance of Time Spent on Web Pages

WEBKDD August 20, 2006, Philadelphia, USA

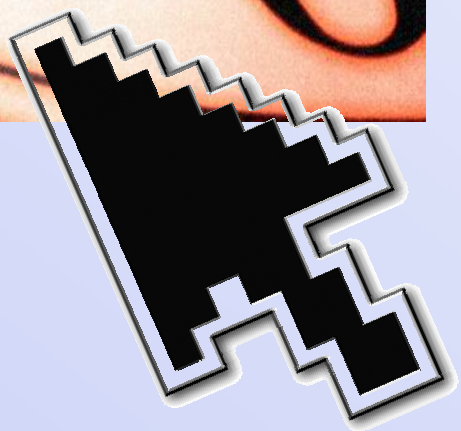
Peter I. Hofgesang

hpi@few.vu.nl



vrije Universiteit

amsterdam



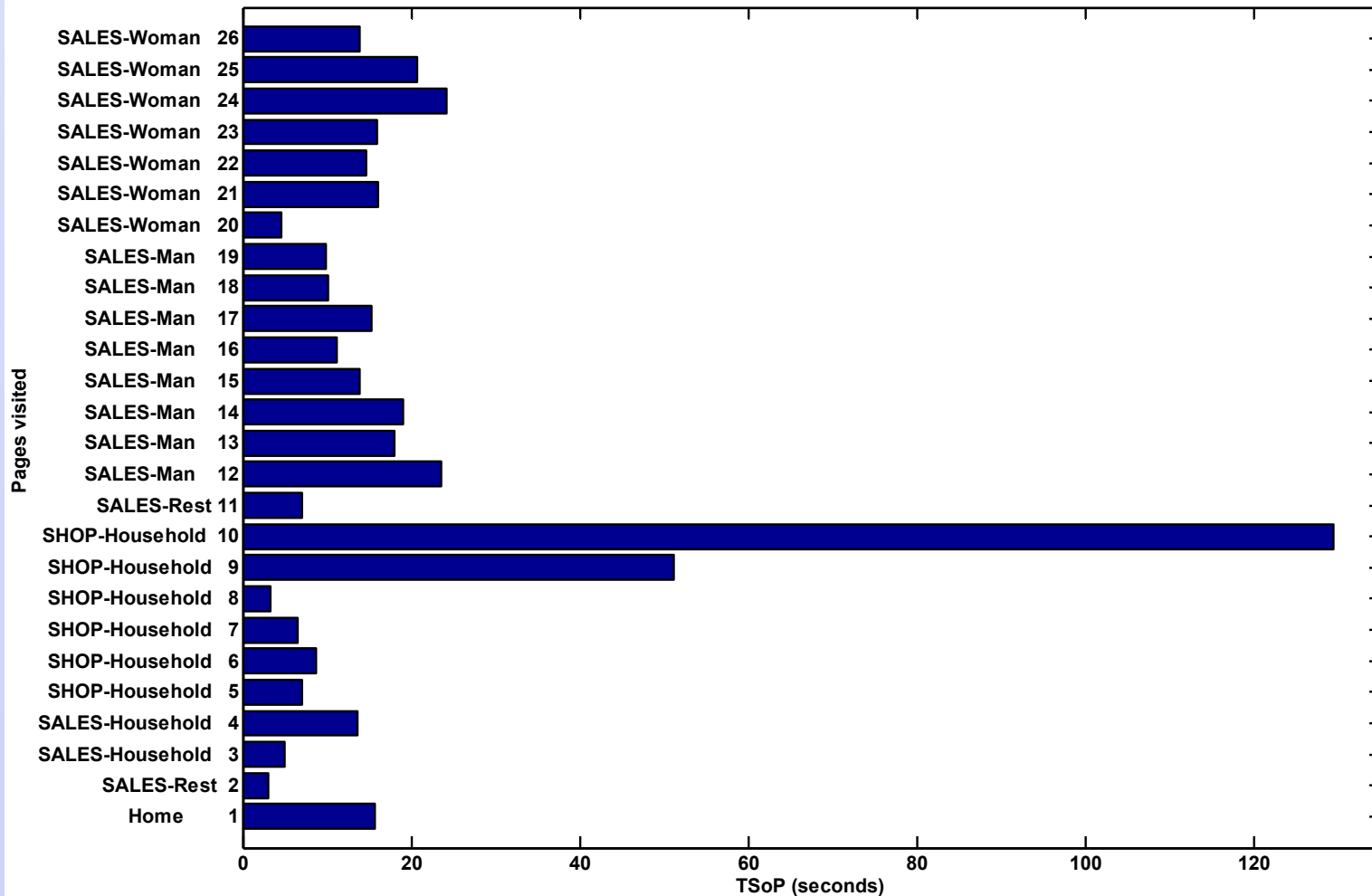
Intention of an online visitor

- Real-world: customers have the ability to **explicitly express** what they are looking for
- Web: **intention is hidden** and can only be partially revealed from implicit indicators in the traces users leave behind

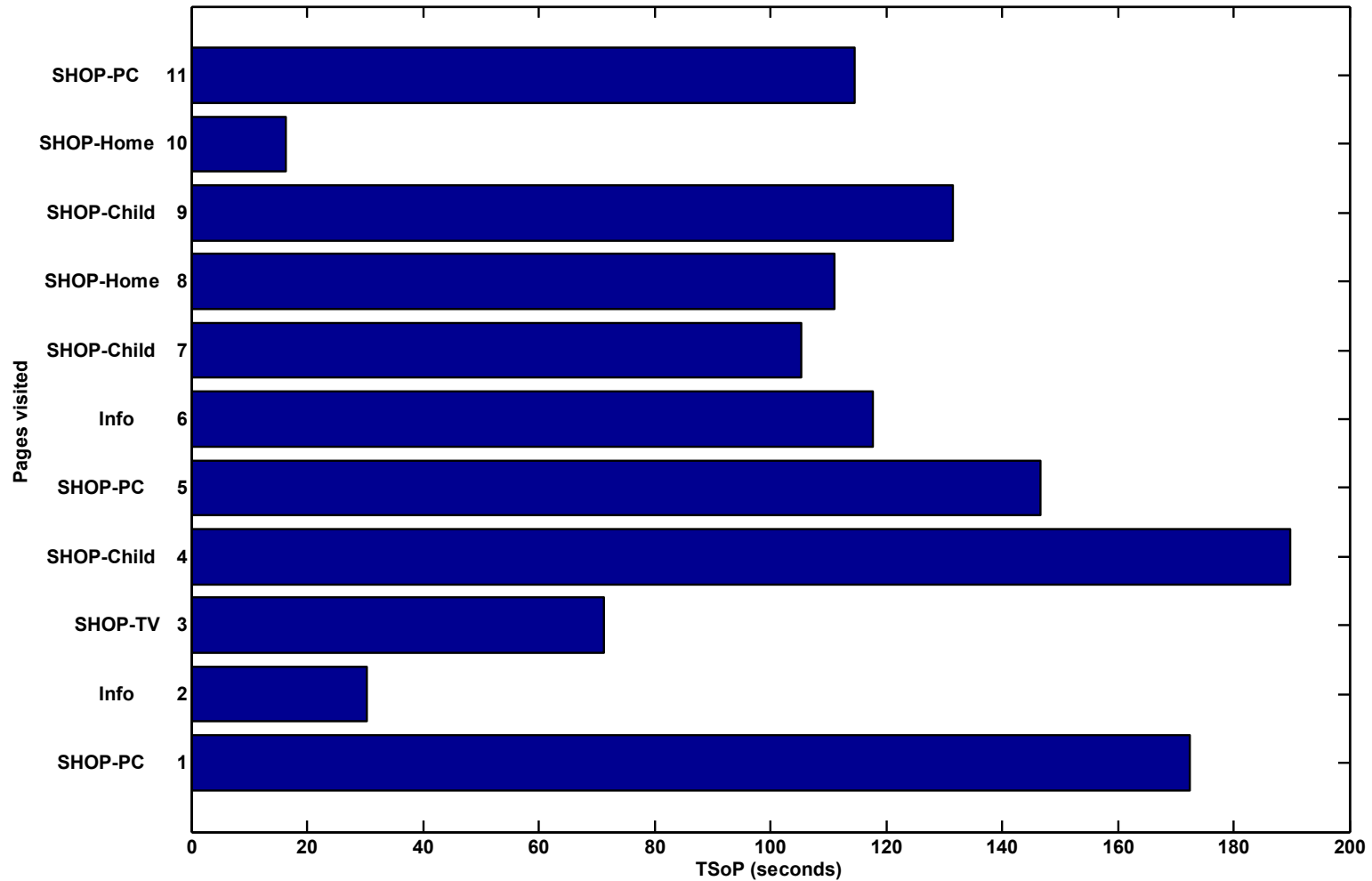
(Broadly) Available information

- **Order** of visited pages (P1 → P2 → P3 ...)
- Page **popularity** (nr. of times visited)
- **Time Spent on Page** (TSP)?
 - claimed to be important in IR, HCI, E-learning
 - only rarely used in WUM
 - details are often not reported(however, preprocessing is not obvious!)

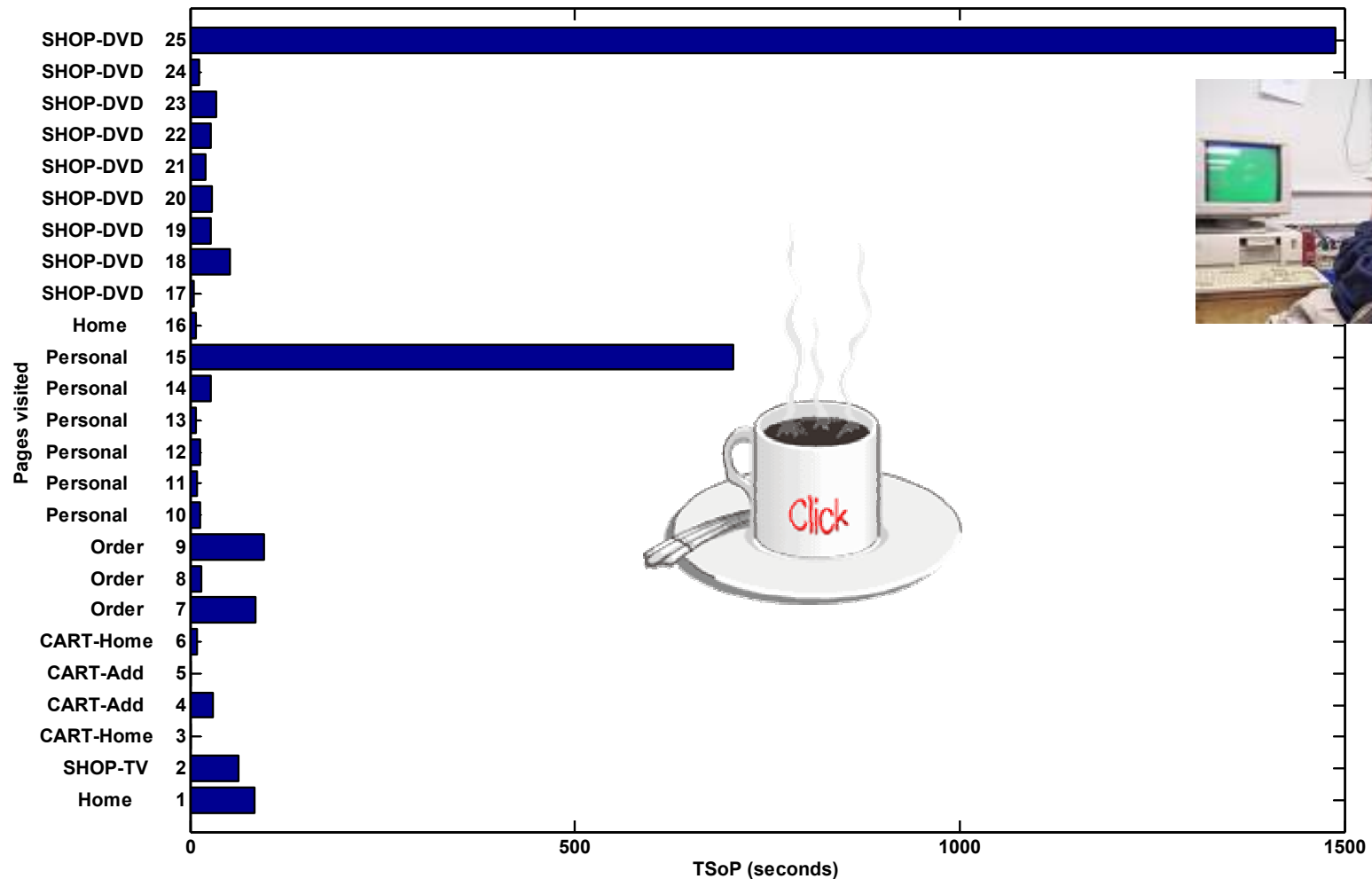
Example I



Example II



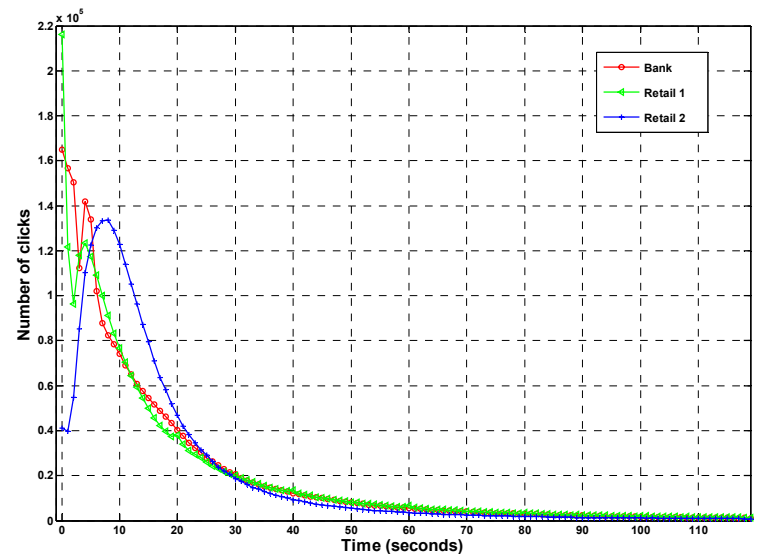
Example III



Influential factors I

$$\mathbf{TSP}_1 = \mathbf{T2} - \mathbf{T1} \quad (\text{optimistic!})$$

- Data preprocessing
 - filtering out robot transactions
 - session identification
- Distraction

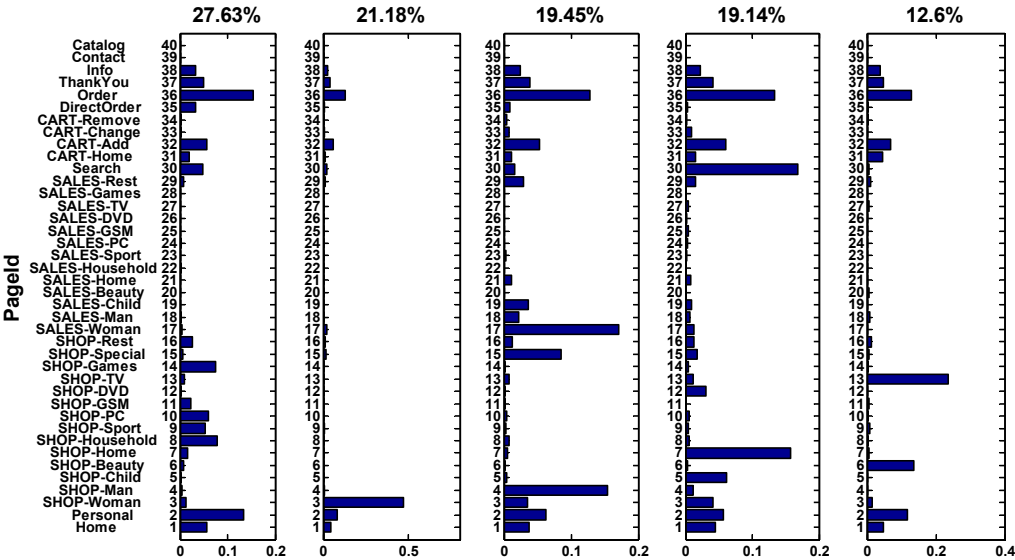


Influential factors II

- Page type (Granularity of pages)
- Hierarchy
- Network bandwidth and server load
- Speed of reading, etc.

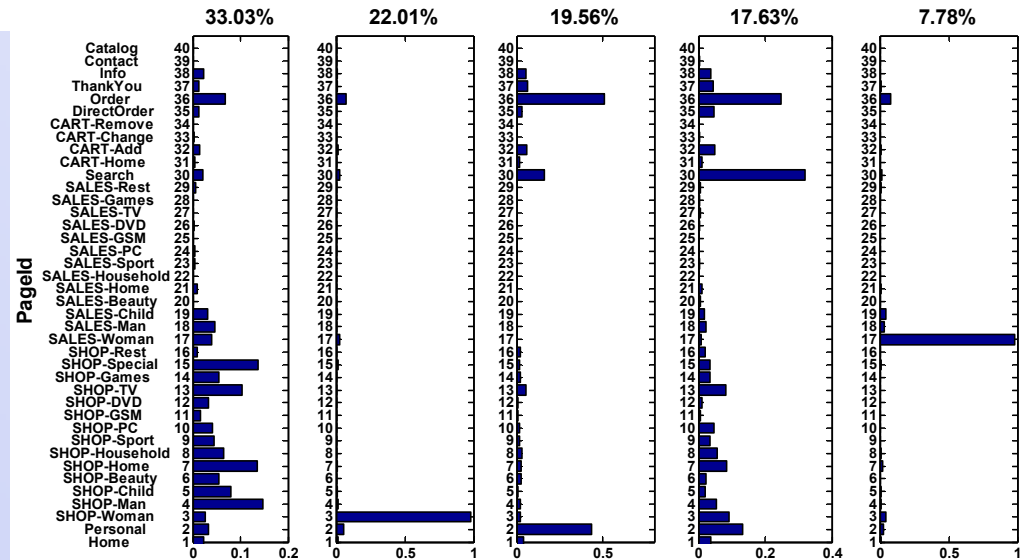
$$TSP_2 = T2 - T1 - T_{networkTraffic} - T_{serverPageGeneration} - T_{distraction}$$

Clustering



Cadez et al. (2001)

Similarity based



Conclusion

- TSP is a **sensitive measure**
- Web log data preprocessing and Time normalization required
- **Added value** in identifying user intention
- For many applications the combination of TSP and frequency may be the optimal choice

Future (current) work

- Objective measures of relevance
- Normally field experiment to provide some kind of labeled data
- **Special testbed** →
 - e.g., in case of a retail shop environment we have special labels for buyers
 - the purchased items indicate user interest and can be compared with the visit

Questions?