Social Computing at MSRA

Natural Language Computing, Microsoft Research Asia

Chin-Yew LIN cyl@microsoft.com

What is Social Computing?

- Social
  - Living together in communities
    - Live ⇨ dynamic
    - Together ⇨ more than one person
    - Community ⇨ same locality
- Computing
  - To determine by the use of a computer
- Social Computing
  - ***To live together in communities using computers***
  - Creating social contexts online via the use of technology

* American Heritage Dictionary

\+ http://www.microsoft.com/sharepoint/capabilities/collaboration/social.mspx

Microsoft Research

Why Social Computing?

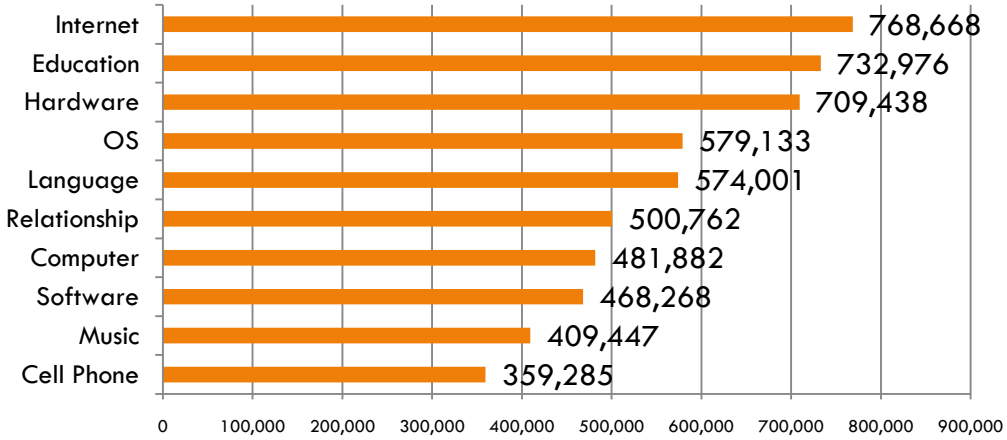
- Web is increasingly social
  - Wikipedia, blogs, Digg, Flickr, del.icio.us, Amazon, Yahoo! Answers, Live QnA, forums, Facebook, MySpace, LinkedIn, …
- Social web
  - People + Content + Network
- How do we turn ***social web*** into ***value*** for ***people***?
  - Assimilation of knowledge → systems pull from data
  - Dissimilation of knowledge → systems to people
  - Elicitation of knowledge → systems pull from people
- "***Pay attention to that man behind the curtain!***"

Microsoft Research

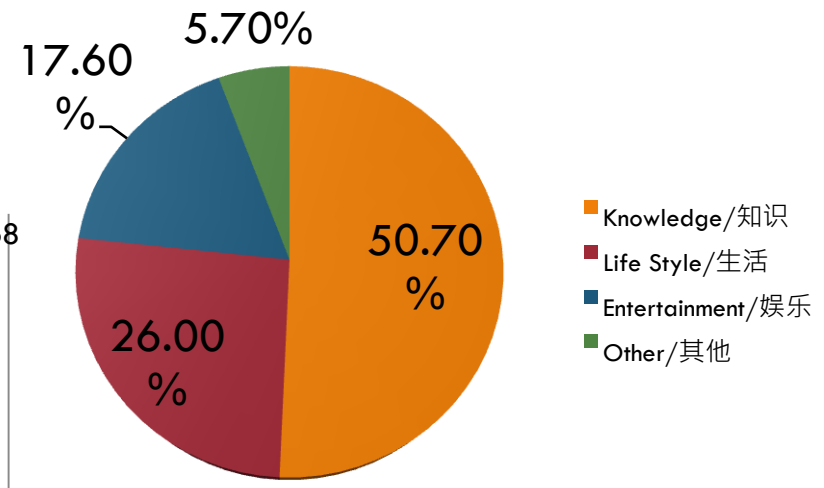
Baidu Zhidao (百度知道)

- 17,012,767 resolved questions in two years' operation
- 8,921,610 are knowledge related
- 96.7% of questions are resolved
- 10,000,000 daily visitors
- 71,308 new questions per day
- 3.14 answers per question

Baidu Zhidao Top 10 Question Types



Baidu Zhidao Question Types Distribution



- http://www.searchlab.com.cn (中国人搜索行为研究/User Research Lab of Chinese Search)

Microsoft Research

Stickiness of Baidu Zhidao

- 据正望咨询调查，“百度知道”跟搜索的关系非常紧密，而且对搜索黏性的提高有很大帮助，根据其统计，“百度知道”已成为百度的一个核心产品。“**百度的用户中有50%搜索‘知道’，其用户量已经超过百度贴吧，与其MP3搜索可相提并论。**”。
- 50% of Baidu users search Baidu Zhidao
- Zhidao search traffic comparable to MP3 search

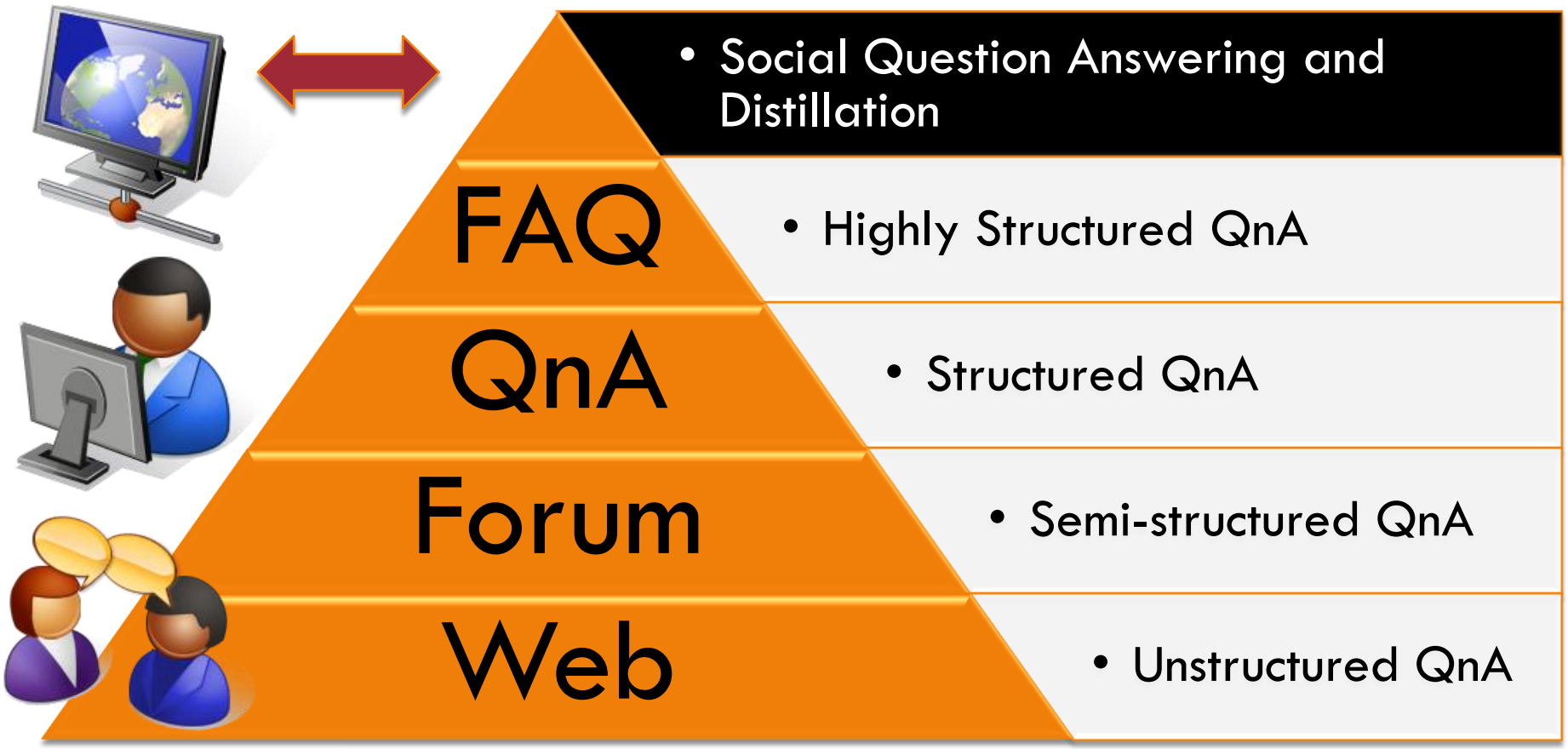
(http://news.csdn.net/n/20080425/115453.html; 04/25/2008)

Microsoft Research

Knowledge Distillation & Dissemination

Social Question Answering

Knowledge Distillation and Dissemination



Microsoft Research

GeoLife 2.0

Building Social Networks Using Human Location History

Yu Zheng, Xing Xie and Wei-Ying Ma

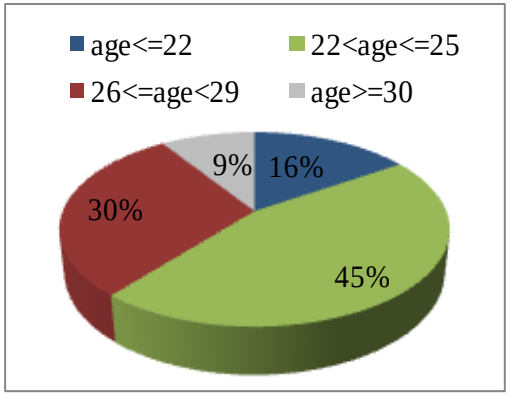
WSM, Microsoft Research Asia

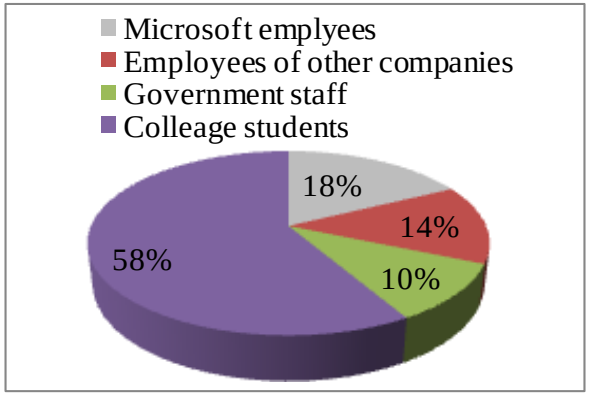
What is GeoLife 2.0?

- A ***location-history-based GPS-data-driven social-networking*** service
  - Enables people to build connections using their GPS trajectories
  - Understand a user and a location
  - Explore the **similarity between users** and the **correlation among locations**

GPS Devices and Users

- 60 devices and 138 users
- From May 2007 to present





Microsoft Research

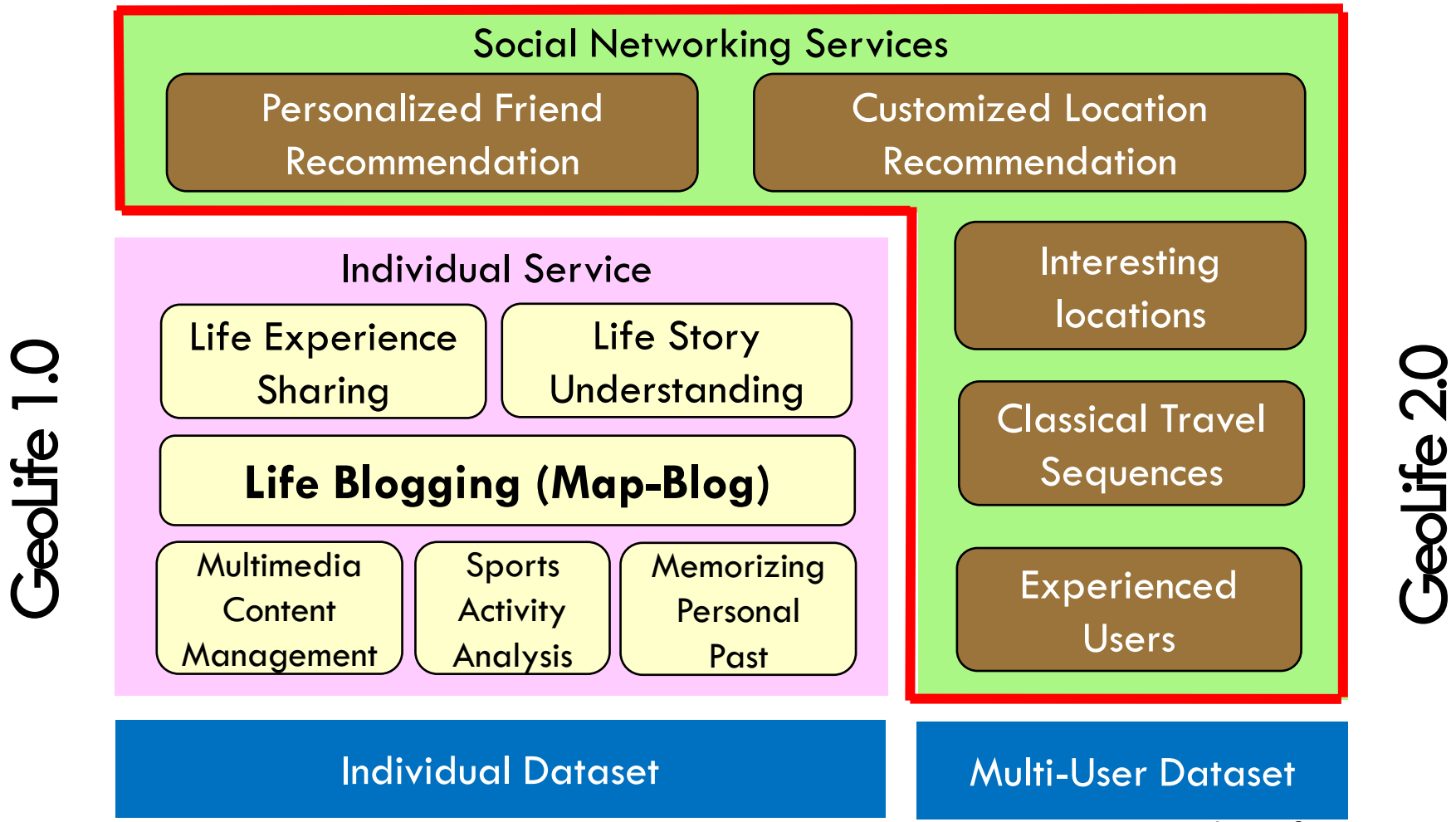
A Large-Scale GPS Data Set

- 10+ million GPS points
- 260+ million kilometers
- 36 cities in China and a few cities in the USA, Korea and Japan

* Bikely: http://www.bikely.com
* GPS Track Route Exchange Forum: http://www.gpsxchange.com
* GPS Sharing: Http://gpssharing.com

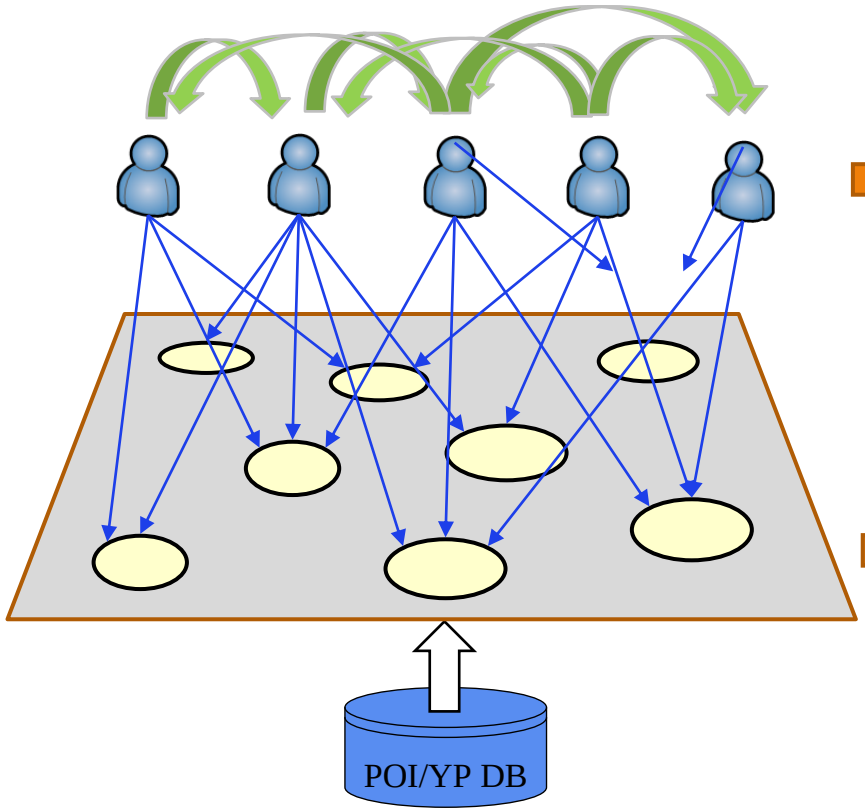
Microsoft Research

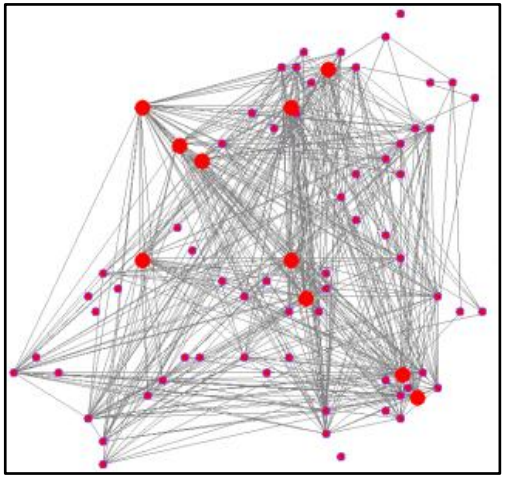
GeoLife Application Architecture

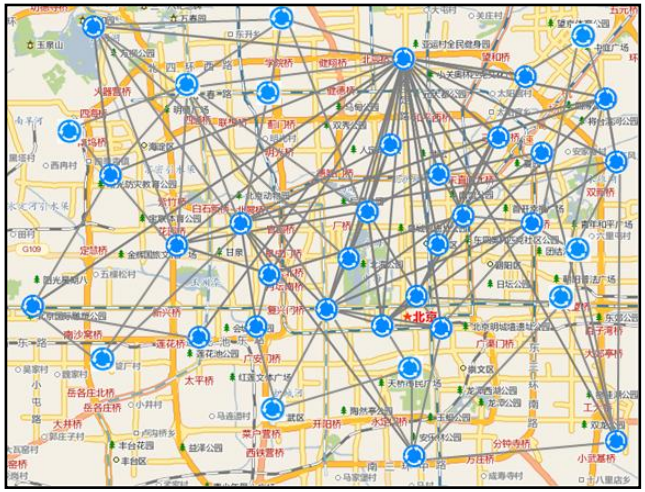


Microsoft Research

A Location-History-Based Social Network

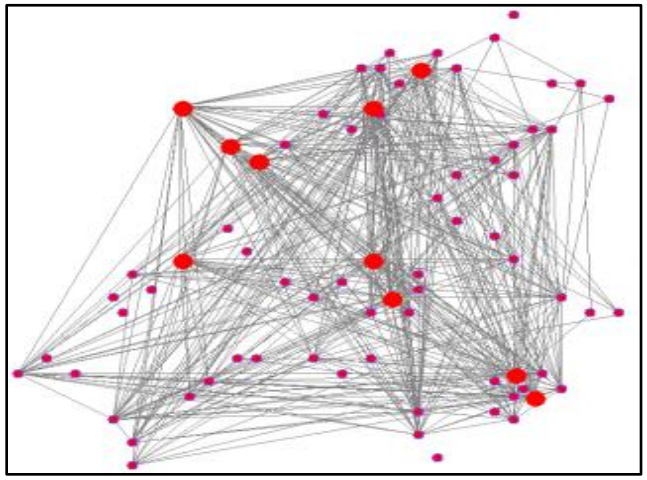


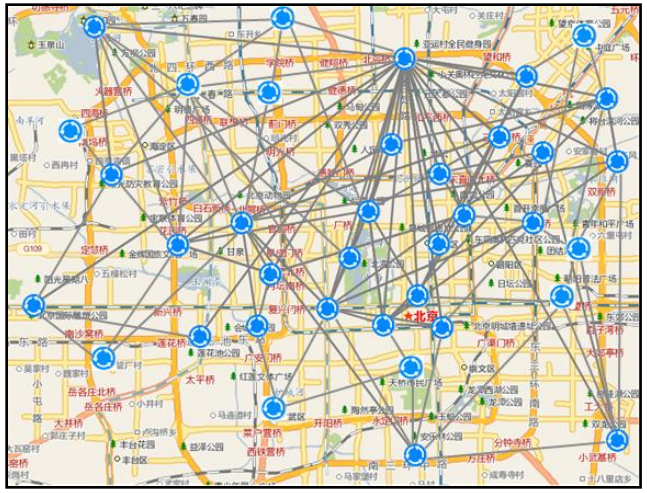




Microsoft Research

Applications





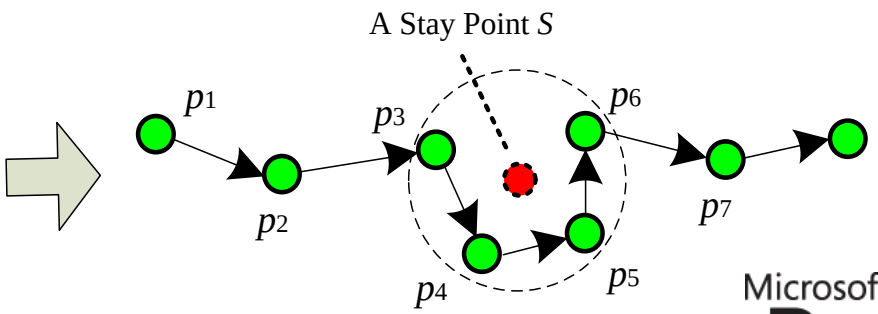
- Understanding People
  - Similar users: Friend recommendation
  - Experienced users: Travel experts recommendation
  - Group users: Community discovery
- Understanding Locations
  - Personalized location recommendation
  - Mining interesting locations
  - Detecting classical travel sequences

Preliminary

- **GPS logs** P **and GPS trajectory**
- **Stay points** $S=\{s_1, s_2, \ldots, s_n\}$.
  - Stands for a geo-region where a user has stayed for a while
  - Carry a semantic meaning beyond a raw GPS point
- **Location history:**
  - Represented by a sequence of stay points with transition intervals

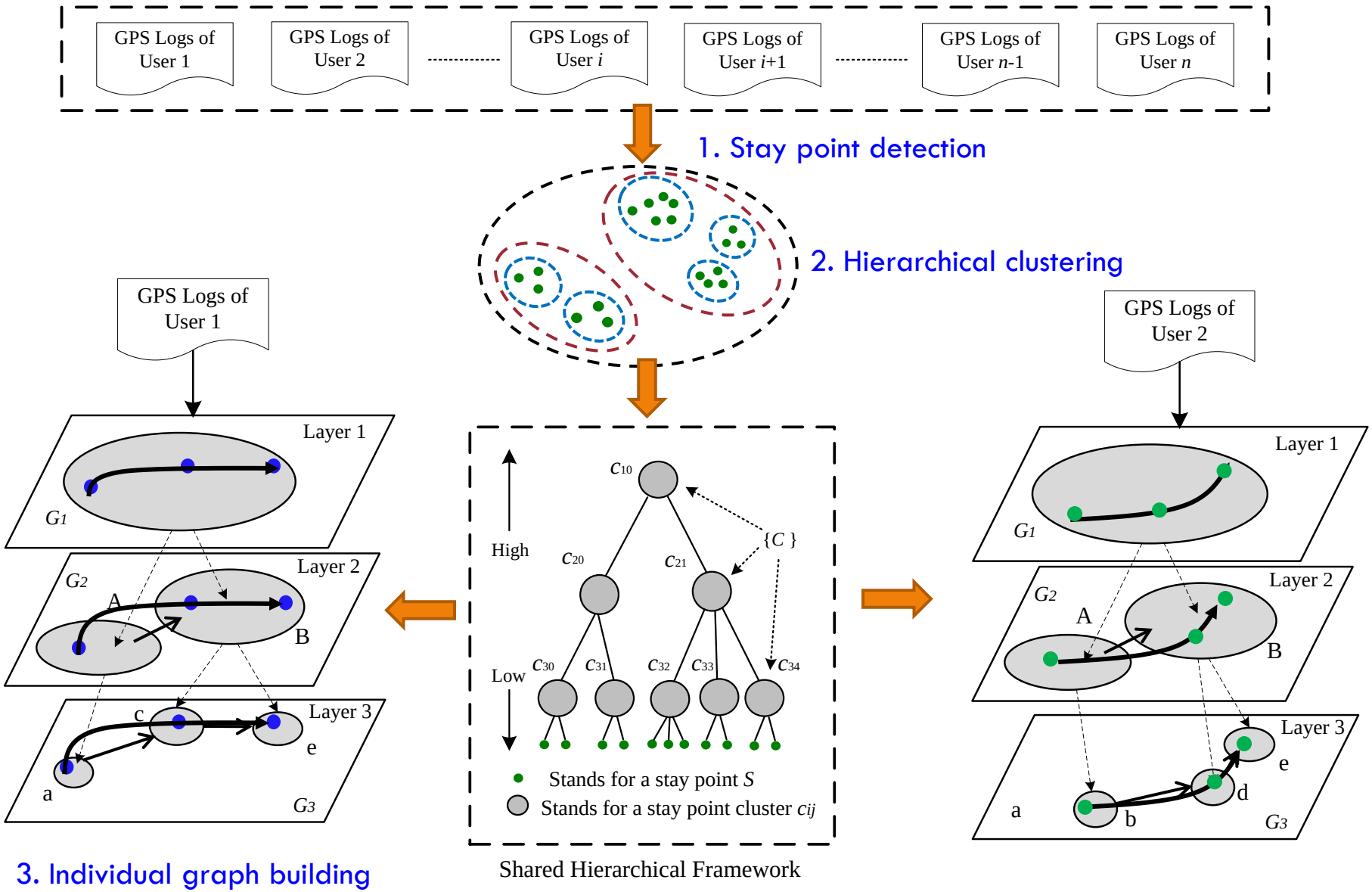
$$LocH = (s_1 \xrightarrow{\Delta t_1} s_2 \xrightarrow{\Delta t_2}, \ldots, \xrightarrow{\Delta t_{n-1}} s_n)$$

	Latitude,	Longitude,	Time
p_1:	Lat1,	Lngt1,	T1
p_2:	Lat2,	Lngt2,	T2
	……...		
p_n:	Latn,	Lngtn,	Tn

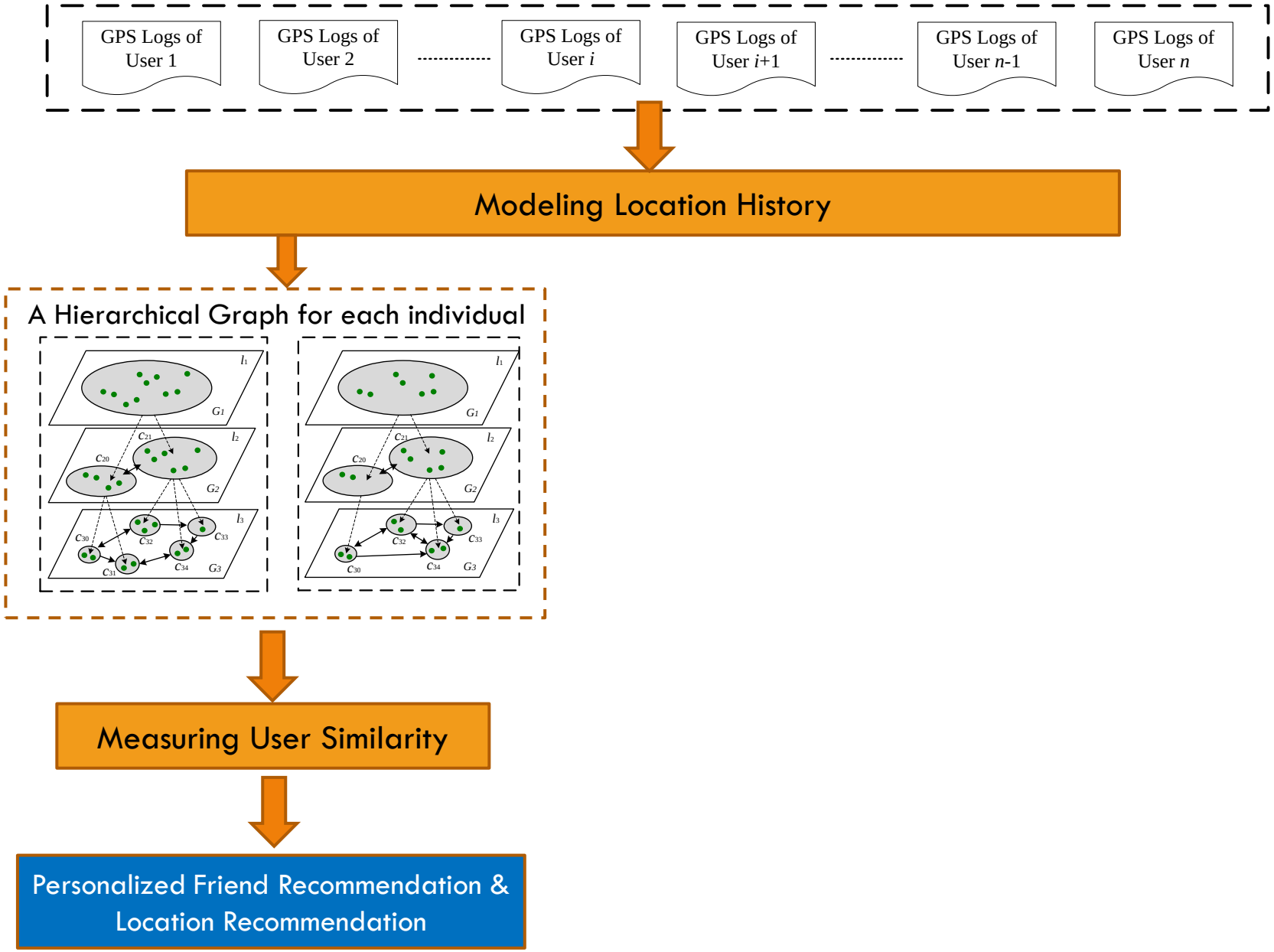


Microsoft Research

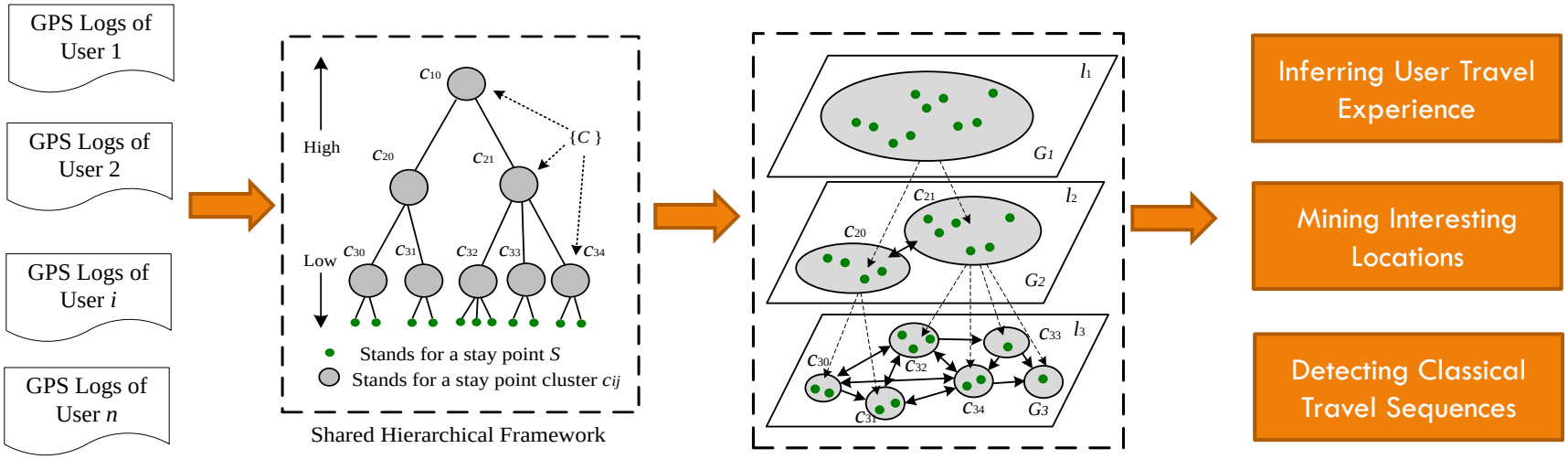
Modeling Individual Location History



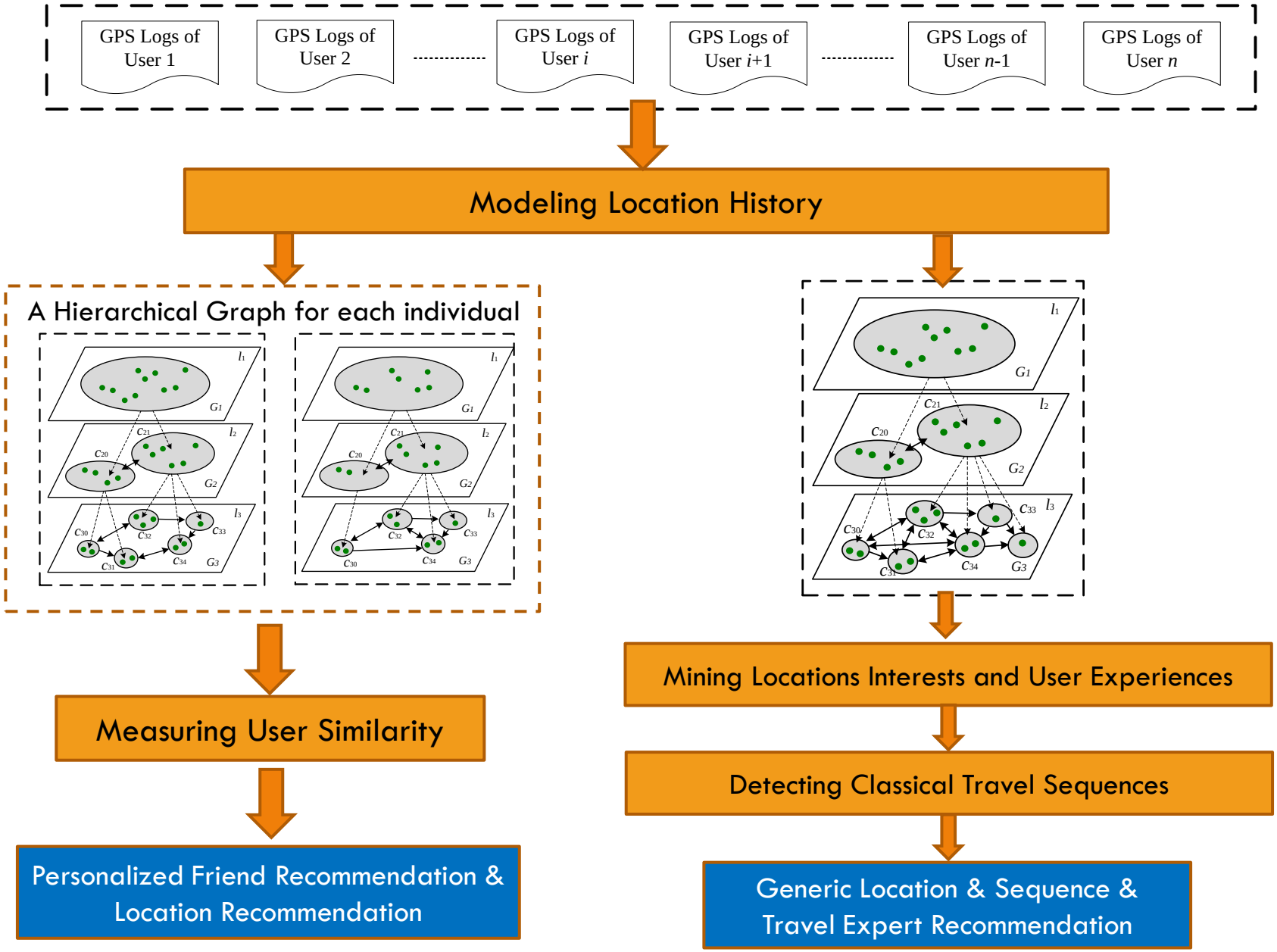
Architecture (1)



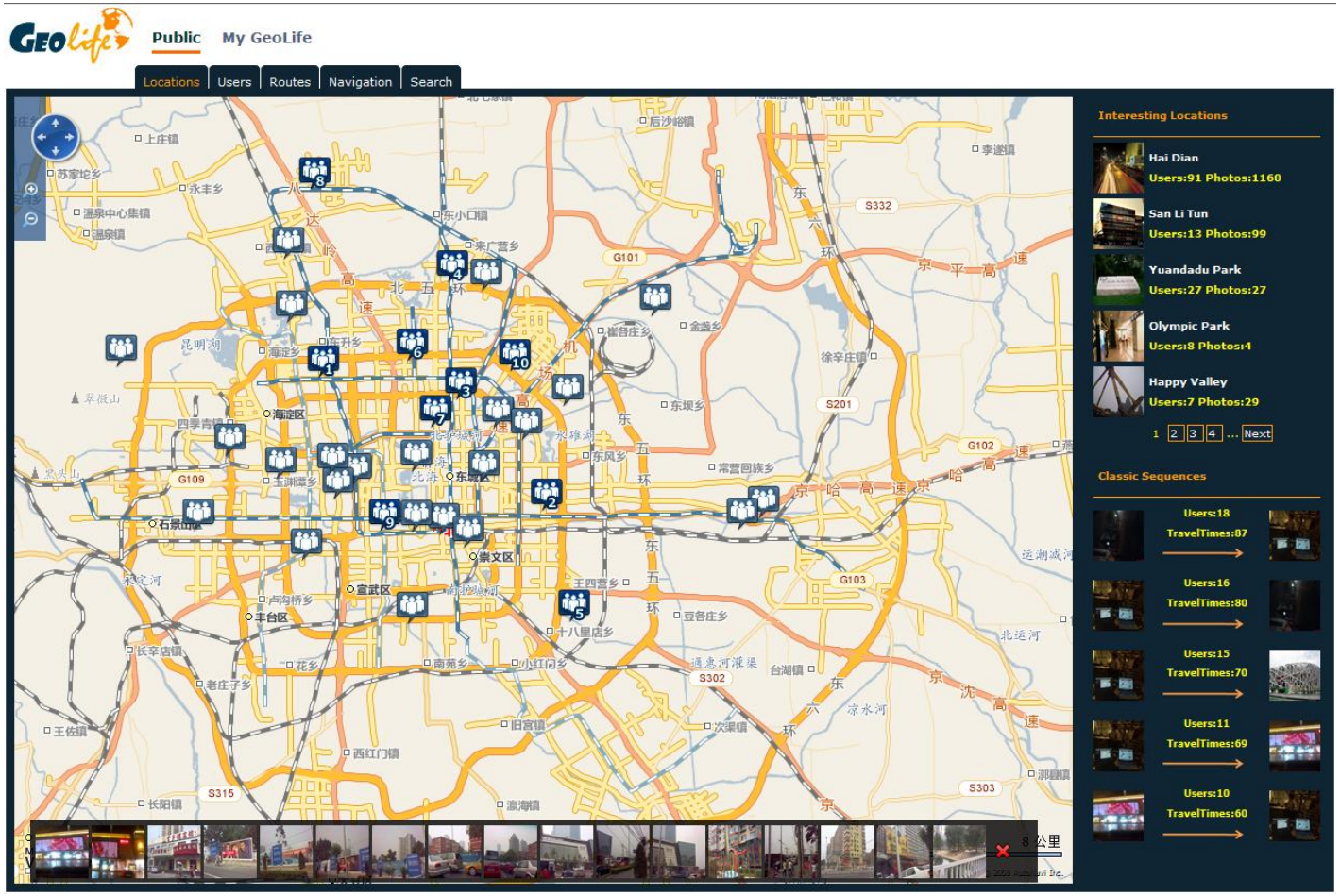
Modeling Multiple Users' Location Histories



Architecture (1)



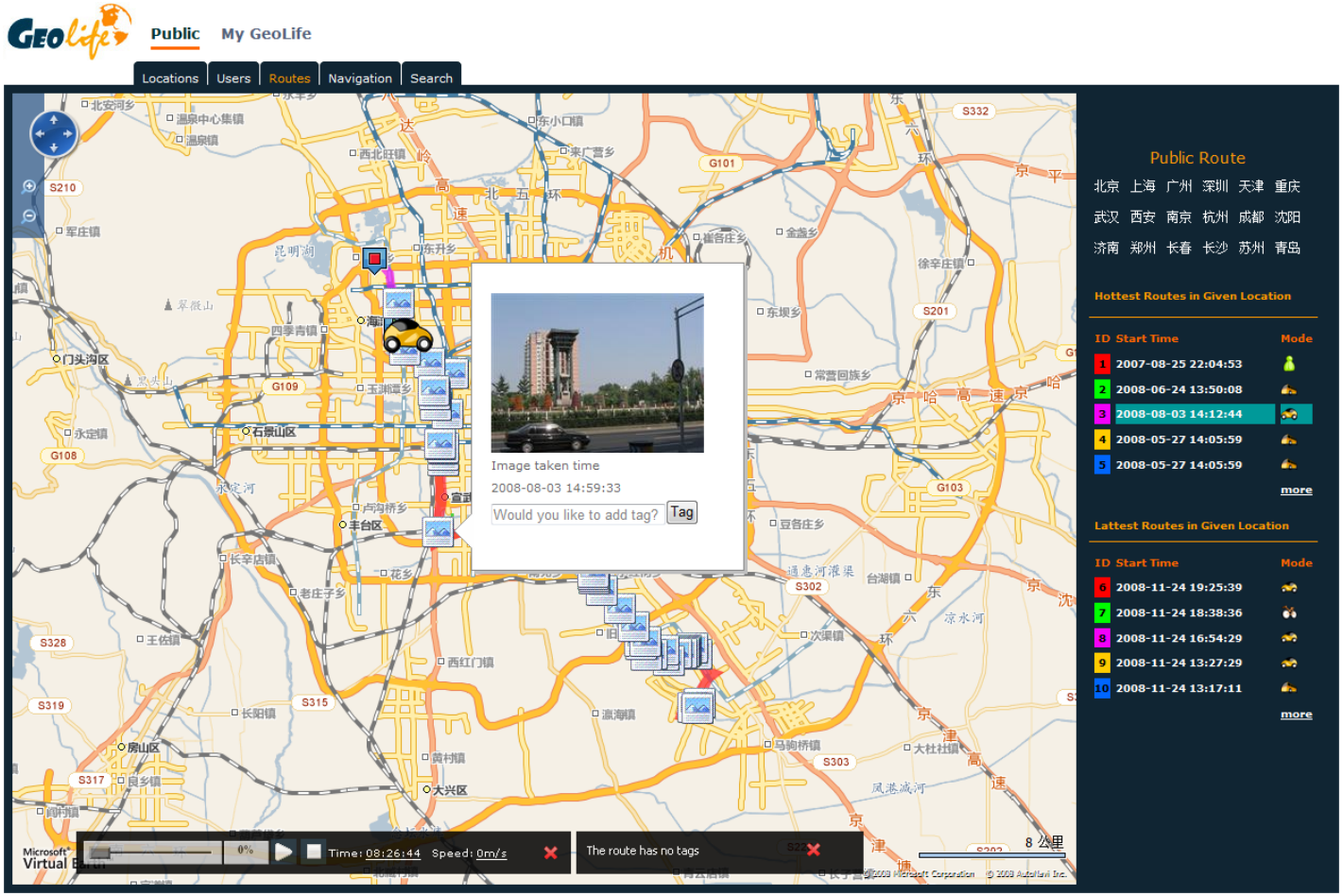
GeoLife: Point of Interest



http://research.microsoft.com/en-us/projects/geolife/

Microsoft Research

GeoLife: Travel Route



Microsoft Research

Summary

- Social web
  - People + Content + Network
- Turn ***social web*** into ***value*** for ***people***
  - Assimilation of knowledge
  - Dissimilation of knowledge
  - Elicitation of knowledge

- "***Pay attention to that man behind the curtain!***"

Microsoft Research

Reference

- Yu Zheng, Lizhu Zhang, Xing Xie and Wei-Ying Ma, Mining Interesting Locations and Travel Sequences from GPS Trajectories, WWW 2009.
- Yu Zheng, Yukun Chen, Xing Xie and Wei-Ying Ma, GeoLife 2.0: A Location-Based Social Networking Service, MDM 2009 (Demo).
- Yu Zheng, Like Liu, Longhao Wang and Xing Xie, Learning Transportation Mode from Raw GPS Data for Geographic Application on the Web, WWW 2008.
- Quannan Li, Yu Zheng, Xing Xie, Yunkun Chen, Wenyu Liu and Wei-Ying Ma, Mining User Similarity Based on Location History, ACM GIS 2008.
- Yu Zheng, Quannan li, Yukun Chen, Xing Xie and Wei-Ying Ma, Understanding Mobility Based on GPS Data, ACM UbiComp 2008.

Microsoft Research