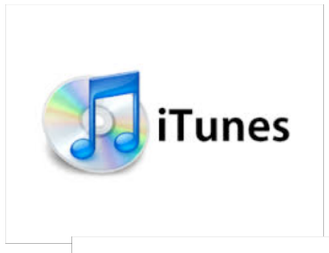
RUTGERS
School of Arts and Sciences

SEMANTIC MODELING, INTEGRATION, AND EPISODIC ORGANIZATION OF PERSONAL DIGITAL TRACES

***Varvara Kalokyri**, Alexander Borgida, Amelie Marian*
Department of Computer Science

PERSONAL DIGITAL TRACES (PDTS) – FRAGMENTED, HETEROGENEOUS



WHAT WAS THE NAME OF THAT RESTAURANT…

- where I went with Mary?
- where we had dinner?
- where we went six months ago?

Some Sources of helpful data

- "with Mary": calendar, email, text (Facebook/Messenger)
- "restaurant": check-ins (Foursquare/Facebook), cell phone GPS logs
- "restaurant": credit card statements, reservations (OpenTable)

MOTIVATION

- Such a collection of personal digital traces (PDTs) can be useful in:
  - helping the user **recall** forgotten details.
  - enabling users **understand** and **query** their PDTs.
- Need for **integrated view** of the user's activities in a sensible uniform manner.
  - basis to connect entities and events into **autobiographical memories**.

CONTRIBUTIONS SO FAR

- **Integrate personal digital traces** by developing techniques to retrieve, store and index PDTs from various heterogeneous sources - **Personal Extraction Tool**
- **Group personal digital traces** with respect to conceptually coherent **episodes** for common everyday events - **extensible approach - Personal Knowledge Base** ExploreDB '17@SIGMOD/PODS
- **Design of a unified and intuitive formalized conceptual model** to link and represent both **PDTs** and their corresponding **episodes.** ODBASE '17
- **Case study** for evaluating our approach with real user's data.
- **Design of an interactive tool** (mobile application) with narrative views of users' digital memories. CIKM '18

CONTRIBUTIONS SO FAR

- **Integrate personal digital traces** by developing techniques to retrieve, store and index PDTs from various heterogeneous sources - **Personal Extraction Tool**
- **Group personal digital traces** with respect to conceptually coherent **episodes** for common everyday events - **extensible approach - Personal Knowledge Base**
- **Design of a unified and intuitive formalized conceptual model** to link and represent both **PDTs** and their corresponding **episodes**.
- **Case study** for evaluating our approach with real user's data.
- **Design of an interactive tool** (mobile application) with narrative views of users' digital memories

INTEGRATING PERSONAL DIGITAL TRACES

- Create an infrastructure to retrieve and store PDTs.
  - **Gather content** from several online services (via APIs, IMAP)
    - Social data - Facebook, Instagram, Twitter, LinkedIn
    - Geolocation data - Foursquare, Facebook, Instagram
    - Email - Gmail, or any other email
    - Calendars - Google Calendar
    - Personal files - Google Drive, Dropbox
    - Web browsing histories - Chrome, Firefox
- Apply entity resolution – who, where dimension
- Apply time extraction - explicating/disambiguating information

CONTRIBUTIONS SO FAR

- **Integrate personal digital traces** by developing techniques to retrieve, store and index PDTs from various heterogeneous sources - **Personal Extraction Tool**
- **Design of a unified and intuitive formalized conceptual model** to link and represent both **PDTs** and their corresponding **episodes**.
- **Group personal digital traces** with respect to conceptually coherent **episodes** for common everyday events - **extensible approach - Personal Knowledge Base**
- **Case study** for evaluating our approach with real user's data.
- **Design of an interactive tool** (mobile application) with narrative views of users' digital memories - **extensible approach**

BACKGROUND

➤Research in psychology has shown two forms of explicit memory

1. Semantic memory – memory of facts and concepts
2. **Episodic memory** – memory of autobiographical events

➤Natural way to remember past events is by any pertinent contextual information; answers to:

➤ ***What, When, Where, Who, Why, How (w5h)***

➤PDTs are inherently contextual due to various forms of metadata

- **When** an email was sent
- **Who** was involved in a conversation
- **Where** a meeting took place
- **What** a file contains
- **Why** a website was accessed
- **How** the information was recorded

CONCEPTUAL MODELING OF PERSONAL DIGITAL TRACES

```
class DOCUMENT is a ENTITY{
  features:
     size : INT;
properties:
     hasPart: set of ENTITY;
     who: set of PERSON;
     what<hasPart: set of DOCUMENT;
     when: set of TIME;
     where: set of LOCATION;
     why: set of GOAL;
}
```

```
class SEND is a ACTION{
  sender<who: PERSON;
  recipients<who: set of PERSON;
  whenSent<when: TIME, …
  }
```

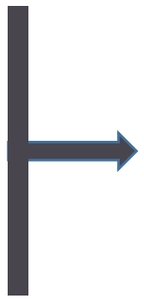
```
class  EMAIL is a DOCUMENT {
  features:
     threadId: STRING;
  properties:
     from<who: PERSON;
     to<who: set of PERSON;
     cc<who: set of PERSON;
     subject < what: TEXT;
     content < what :TEXT;
     attachments < what: set of DOCUMENT;
  actions:
     send: SEND
     reply: REPLY
  constraints:
     from =send.sender;
     send.whenSent<when;   …
  }
```

CONTRIBUTIONS SO FAR

- **Integrate personal digital traces** by developing techniques to retrieve, store and index PDTs from various heterogeneous sources - **Personal Extraction Tool**
- **Design of a unified and intuitive formalized conceptual model** to link and represent both **PDTs** and their corresponding **episodes**.
- **Group personal digital traces** with respect to conceptually coherent **episodes** for common everyday events - **extensible approach** - **Personal Knowledge Base**
- **Case study** for evaluating our approach with real user's data.
- **Design of an interactive tool** (mobile application) with narrative views of users' digital memories

GROUP PDTS INTO COHERENT EPISODES

- Goal: Organize & summarize PTDs into episodes
  - Emails concerning a dinner
  - OpenTable reservation at a restaurant
  - Facebook checkin with photos
  - Credit card payment



Part of the narrative for going out to eat

- To do so, we use a set of higher level ***prototypical plans*** that the user and her/his community frequently engage in.
- **Scripts** : **prototypical plans**, "a predetermined, stereotyped sequence of actions that defines a well-known situation" (Schank & Abelson '77).
- Scripts are composed of sub-scripts, and abstract some of their details.

GROUP PDTS INTO COHERENT EPISODES

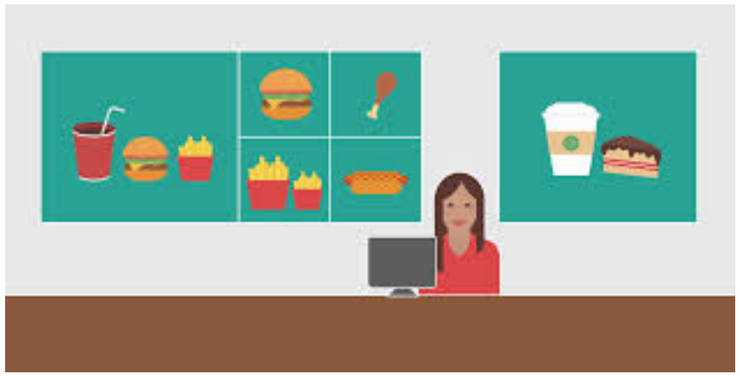
- Example - Going out to eat at a restaurant
  - **Script** would provide description of possible "event flows"

Arrange where & when to go

Make reservation

Call a cab/uber

Go to restaurant



Order food

[….]

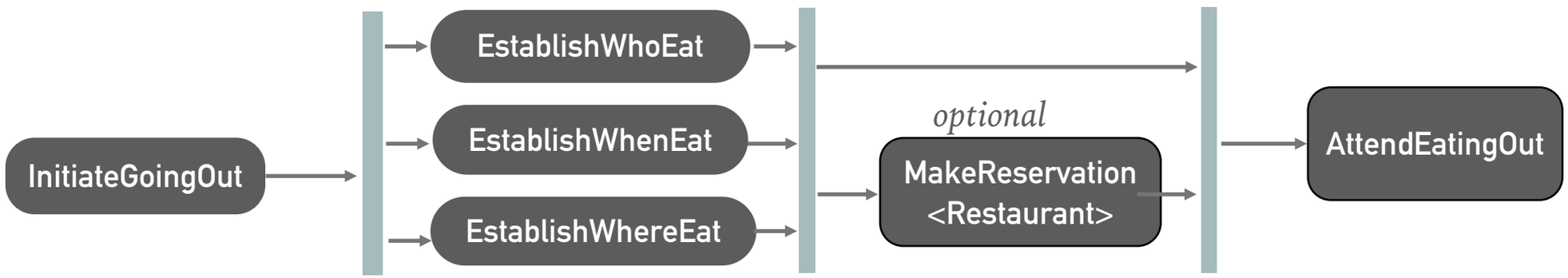
Pay

[….]

ONTOLOGY FOR SCRIPTS & SCRIPT PROPERTIES

```
class Eating_Out is a SCRIPT{
  locals:
    whoAttended <who : set of PERSON
    whereEating < where : EATERY
    whenEating< when :TIME
    whatEaten<what: set of FOODS
    purpose<why : GOAL
  body:
    InitiateGoingOut ;
    (And
      EstablishWhenToEat,
      EstablishWhoWillEat,
      EstablishWhereToEat) ) ;
      (Optional
        MakeRestaurantReservation);
      AttendEatingOut)
}
```

```
class AttendEatingOut is a SCRIPT{
  body:
    GetToEatery ; CheckIn;
    (OR(OrderFood; BeServed), SelfServeFood) ;
      Eat; MakeAPayment; LeaveEatery;
}
```



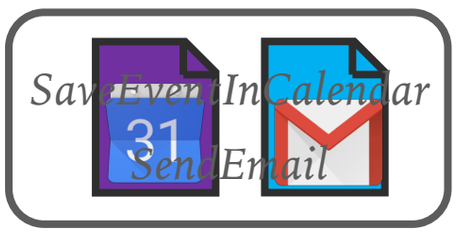
UML activity diagram for Eating_Out

ALGORITHM FOR INSTANTIATING SCRIPT INSTANCES

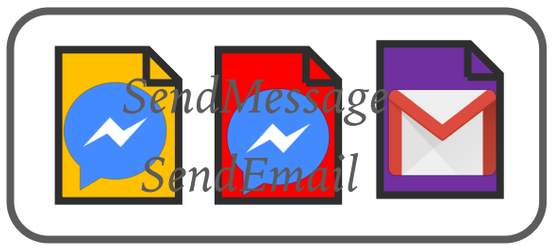
Personal digital traces

Retrieve *relevant PDTs*

Strong evidence



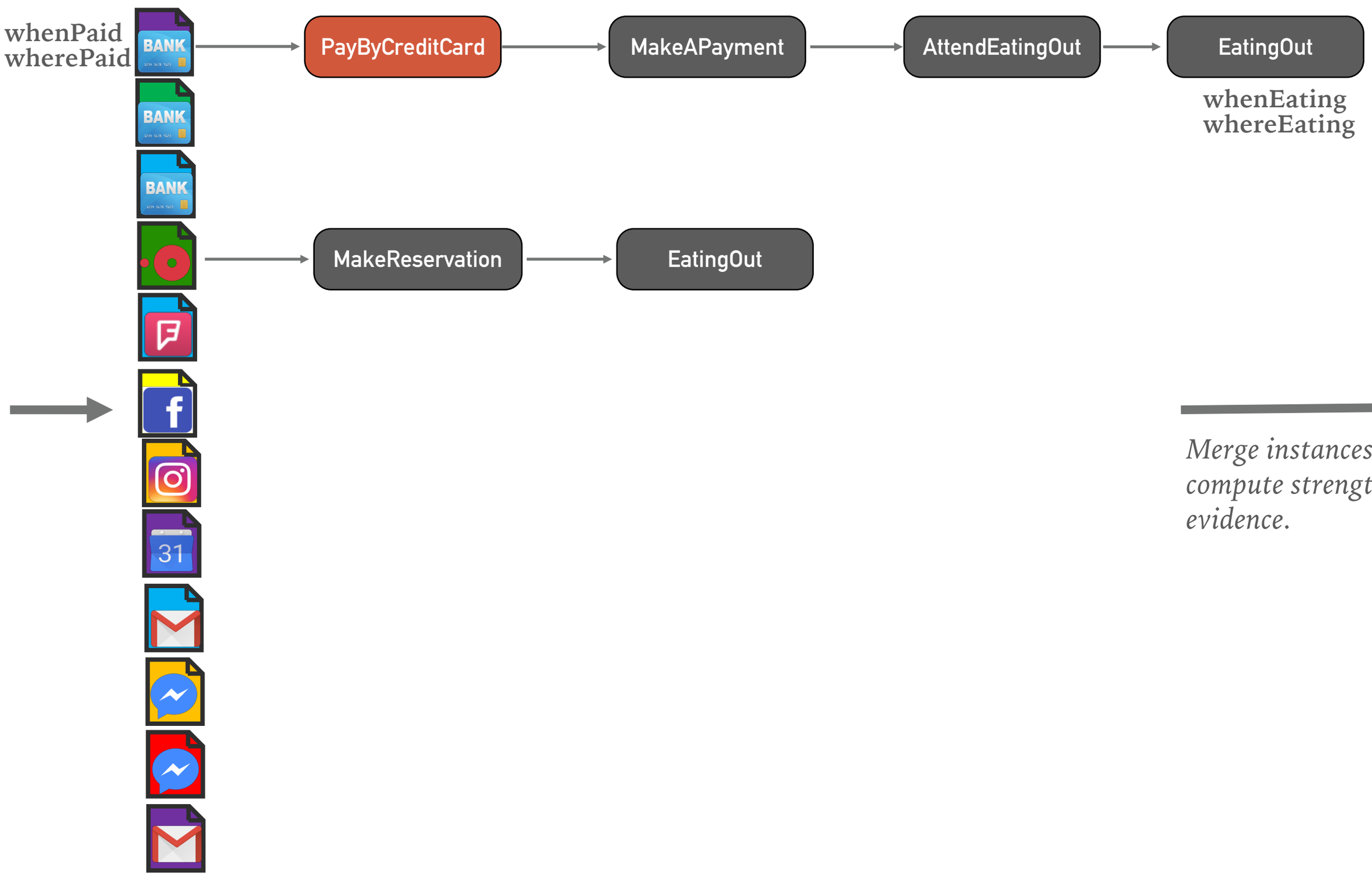
Mild evidence



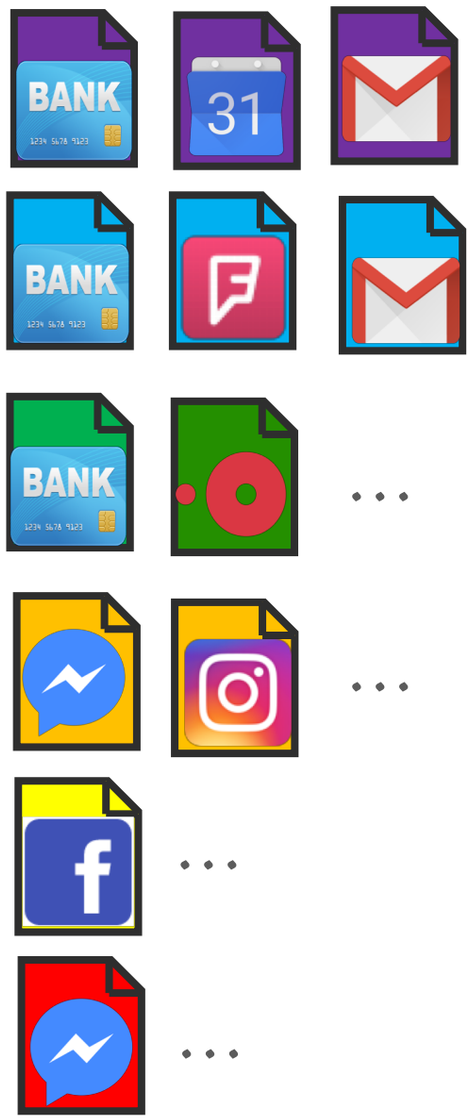
Weak evidence

Create script instance with some information filled in

ALGORITHM FOR INSTANTIATING SCRIPT INSTANCES



ALGORITHM FOR INSTANTIATING SCRIPT INSTANCES



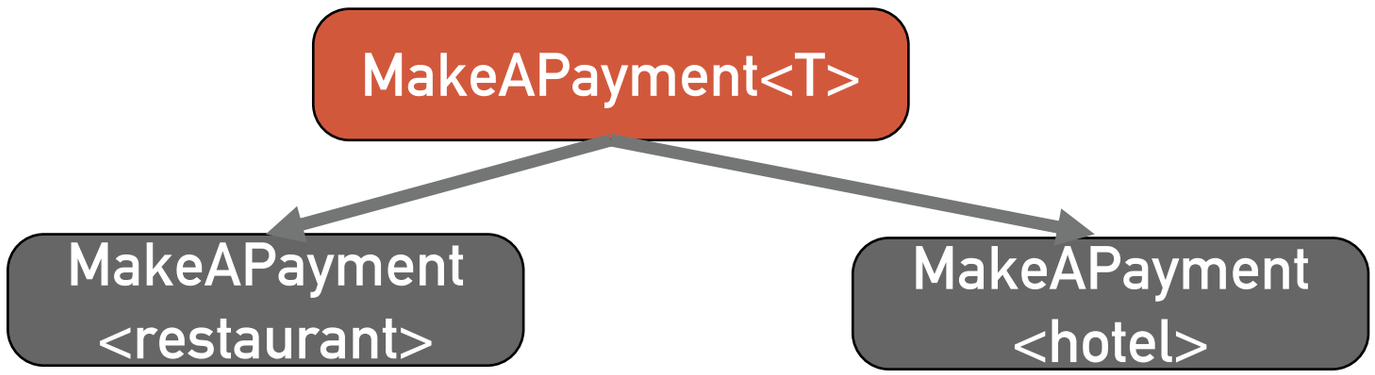
who…
where…
when…

who…
what…
when…
where…

Get more PDTs based on the script body

EXTENSIBLE APPROACH FOR SCRIPT INSTANTIATION

1. Creation of **"trigger words/phrases"**
   - **Verbs** for goal events + synonyms, hyponyms -Wordnet, ConceptNet5
   - **w5h participants** of the goal event - Verbnet, Framenet
2. Declarative definition of **Scripts**
   - Top-level script, subscripts, atomic tasks, locals (w5h info), sequencing, relationship
   - All the scripts/subscripts are **parametric/generic**



3. Declarative Description of **Evidence/Clues** (strong/weak)
4. Declarative definition for **mapping PDTs locals to Script locals**

CONTRIBUTIONS SO FAR

- **Integrate personal digital traces** by developing techniques to retrieve, store and index PDTs from various heterogeneous sources - **Personal Extraction Tool**
- **Design of a unified and intuitive formalized conceptual model** to link and represent both **PDTs** and their corresponding **episodes**.
- **Group personal digital traces** with respect to conceptually coherent **episodes** for common everyday events - **extensible approach - Personal Knowledge Base**
- **Case study** for evaluating our approach with **real** user's data.
- **Design of an interactive tool** (mobile application) with narrative views of users' digital memories

CASE STUDY: EATING OUT

- *Goal*: Find, among users PDTs instances of eating at various restaurants.
- **Three** users: Alice, Bob, Charlie
- **Six-month** period PDTs
- **Four** types of sources:
  - **messaging** (e.g., email, Facebook messenger, Hangouts)
  - **calendaring** (e.g. Google Calendar)
  - **financial transactions** (e.g. bank and credit card statements)
  - **location services** (e.g. Foursquare, Facebook checkins).

GOLDEN SET

- The identification of the golden set a posteriori is difficult.
- Every user went carefully over their recorded PDTs and identified all data that pertained to Eating Out events.

Alice	Bob	Charlie
63	21	(116) 40

Number of identified `Eating_out` events per user

RELEVANT PDTS TO THE EATING_OUT SCRIPT

	Alice	Bob	Charlie
Email/Messaging	**56**	**52**	26
Calendar	**-**	**14**	**7**
Financial Data	44	17	**136(49)**
Location	**9**	-	-

Number of PDTs relevant to the `Eating_Out` script per source per user

EVALUATION METRICS

- **Percentage of events retrieved**

$$\frac{\text{\# relevant instances retrieved}}{\text{\# all } \mathbf{relevant} \text{ instances}}$$

- **Overall Precision**

$$\frac{\text{\# relevant instances retrieved}}{\text{\# all } \mathbf{retrieved} \text{ instances}}$$

- **Precision@k**

EXPERIMENTAL RESULTS – RECALL

	Alice	Bob	Charlie
Email/Messaging	0.59	0.86	0.06 (0.18)
Calendar	-	0.29	0.05 (0.15)
Financial Data	0.67	0.52	0.89 (0.68)
Location	0.14	-	-
Email/Messaging+Financial Data	0.98	1	0.95 (0.85)
Calendar+Financial Data	0.67	0.76	0.95 (0.83)
Location+Financial Data	0.68	0.52	0.89 (0.68)
Calendar+Email/Messaging	0.59	0.86	0.11 (0.33)
Email/Messaging+Location	0.7	0.86	0.06 (0.18)
All sources	1	1	1

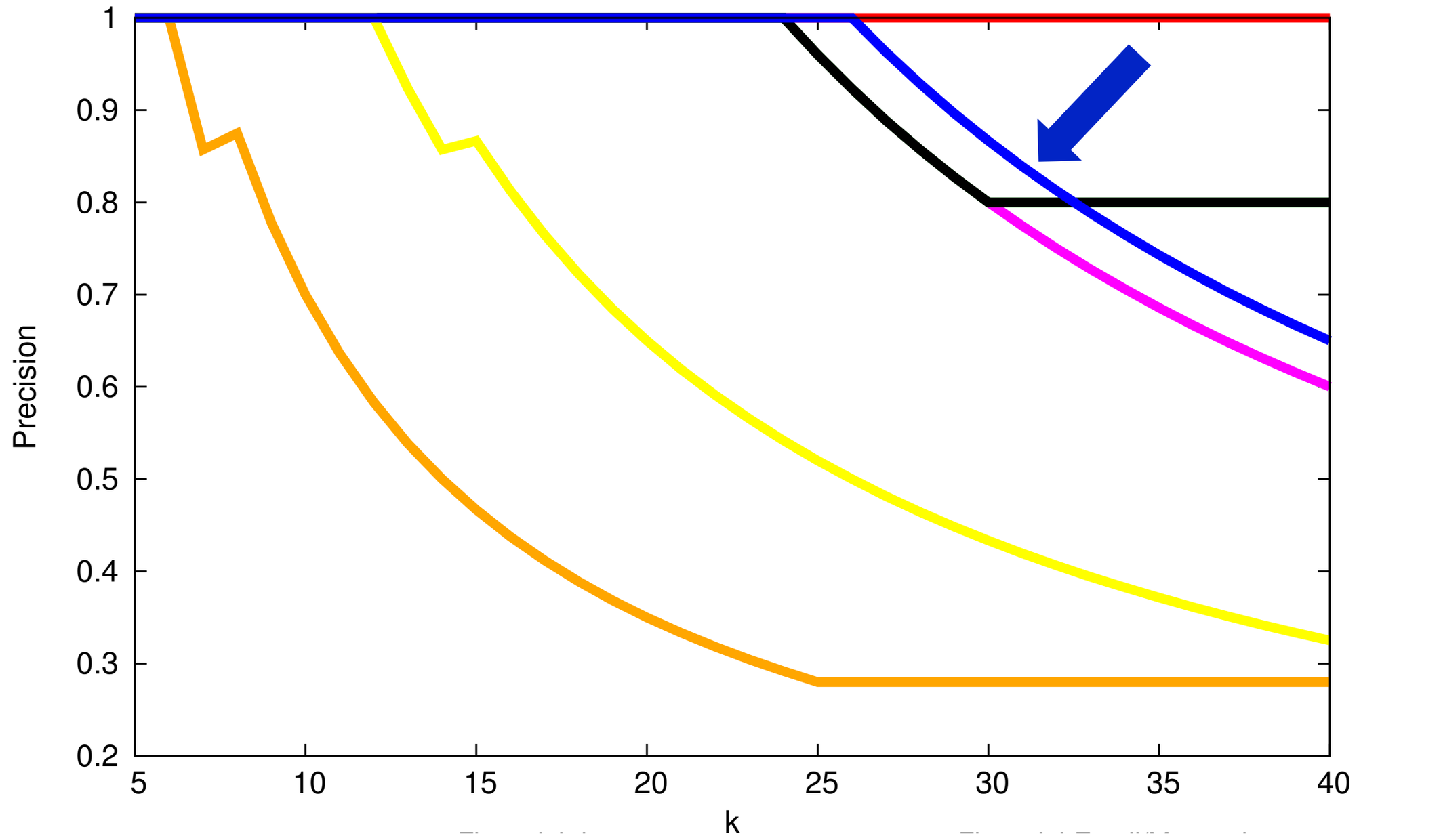
Percentage of Eating_Out episodes retrieved per (set of) sources, per user

EXPERIMENTAL RESULTS – PRECISION

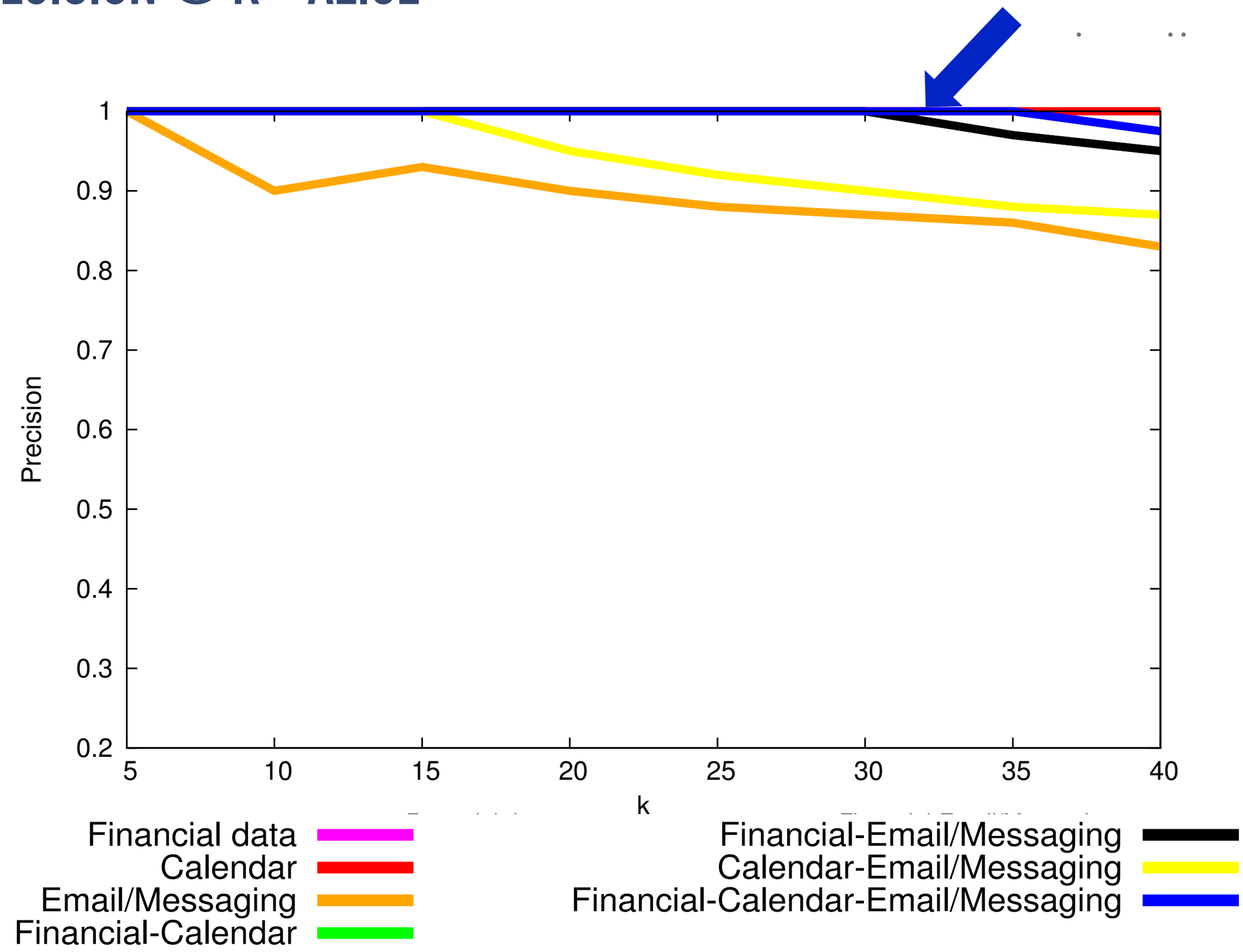
	Alice	Bob	Charlie
Email/Messaging	0.66	0.33	0.33
Calendar	-	0.43	0.86
Financial Data	0.95	0.65	0.82 (0.55)
Location	1	-	-
Email/Messaging+Financial Data	0.75	0.32	0.69 (0.4)
Calendar+Financial Data	0.95	0.52	0.83 (0.59)
Location+Financial Data	0.96	0.65	0.82 (0.55)
Calendar+Email/Messaging	0.66	0.32	0.46
Email/Messaging+Location	0.7	0.35	0.33
All sources	0.75	0.32	0.69 (0.48)

Overal Precision per (set of) sources, per user

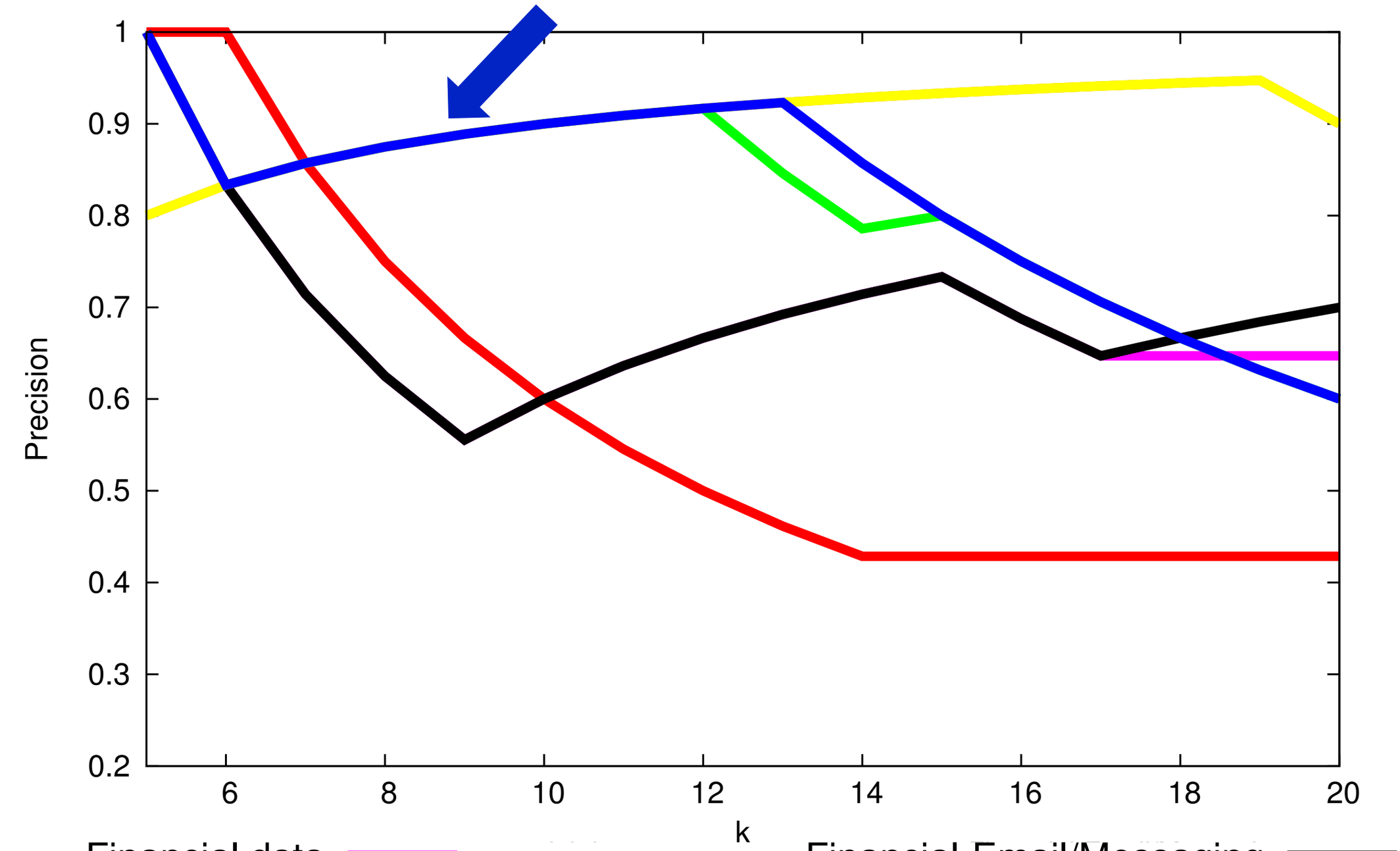
PRECISION @ K – CHARLIE



PRECISION @ K – ALICE



PRECISION @ K – BOB



CONTRIBUTIONS SO FAR

- **Integrate personal digital traces** by developing techniques to retrieve, store and index PDTs from various heterogeneous sources - **Personal Extraction Tool**
- **Design of a unified and intuitive formalized conceptual model** to link and represent both **PDTs** and their corresponding **episodes**.
- **Group personal digital traces** with respect to conceptually coherent **episodes** for common everyday events - **extensible approach - Personal Knowledge Base**
- **Case study** for evaluating our approach with real user's data.
- **Design of an interactive tool** (mobile application) with narrative views of users' digital memories

YOUR DIGITAL SELF

- A mobile-based personal information organization application.
  - To be used to implement and evaluate our research through user studies and surveys.
- Provide users with narrative views of their digital memories.

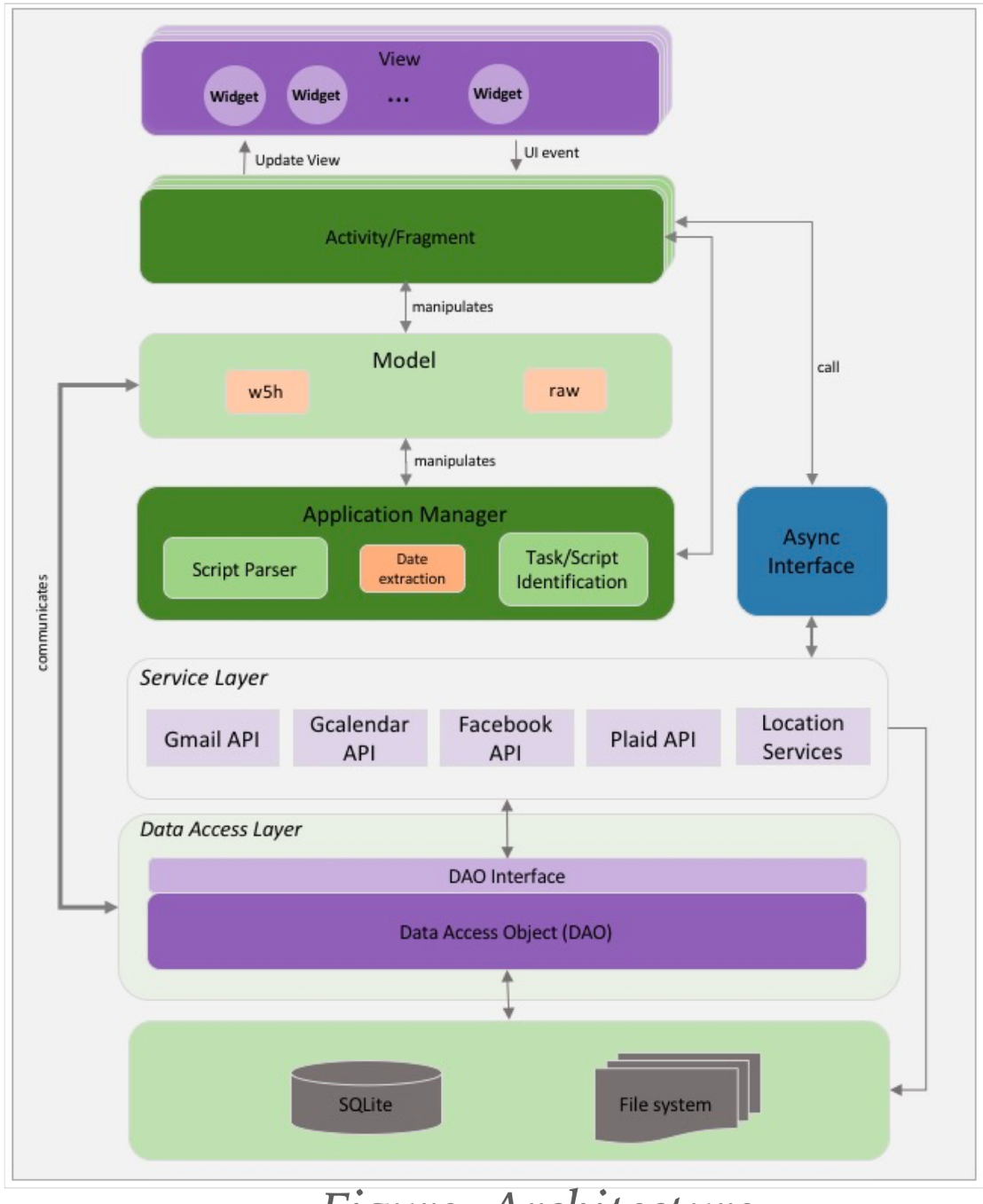
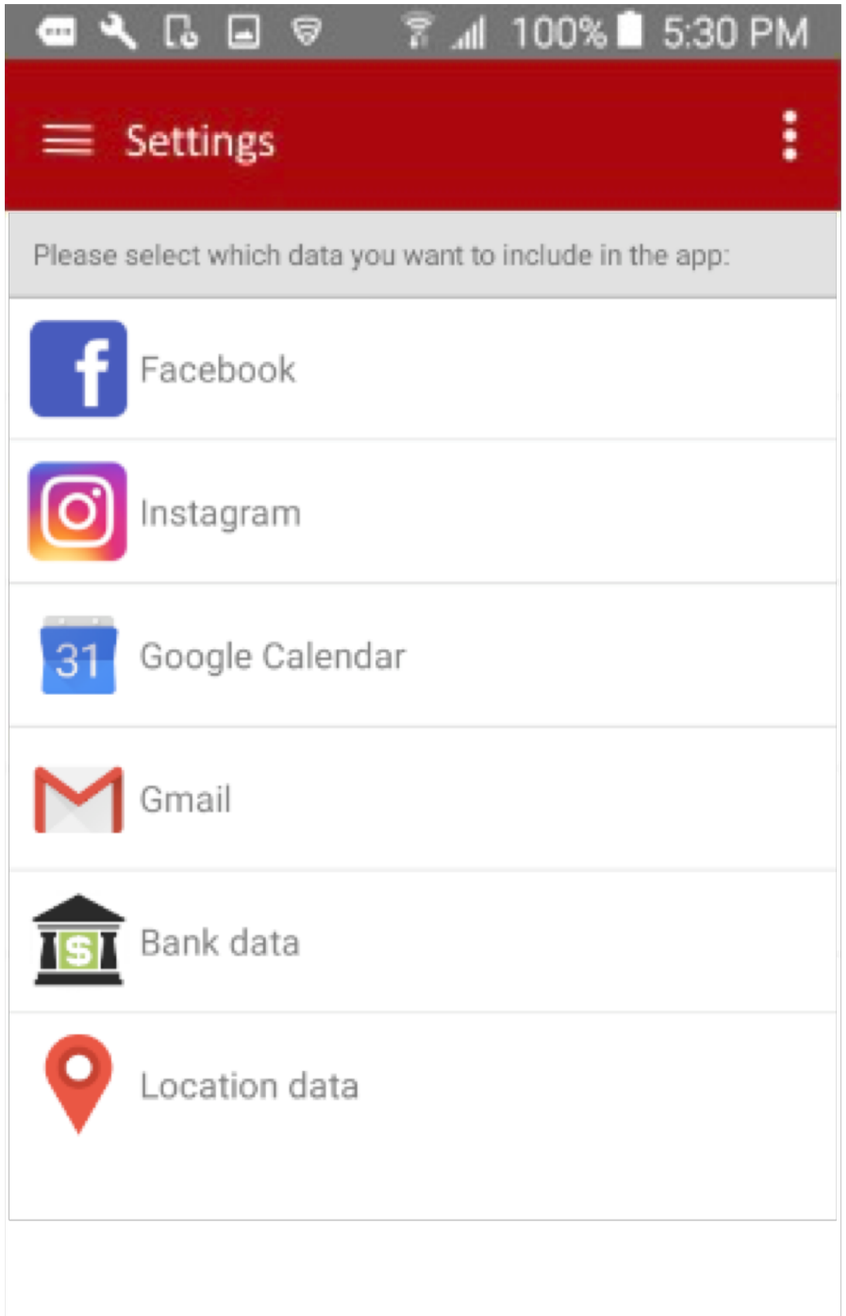
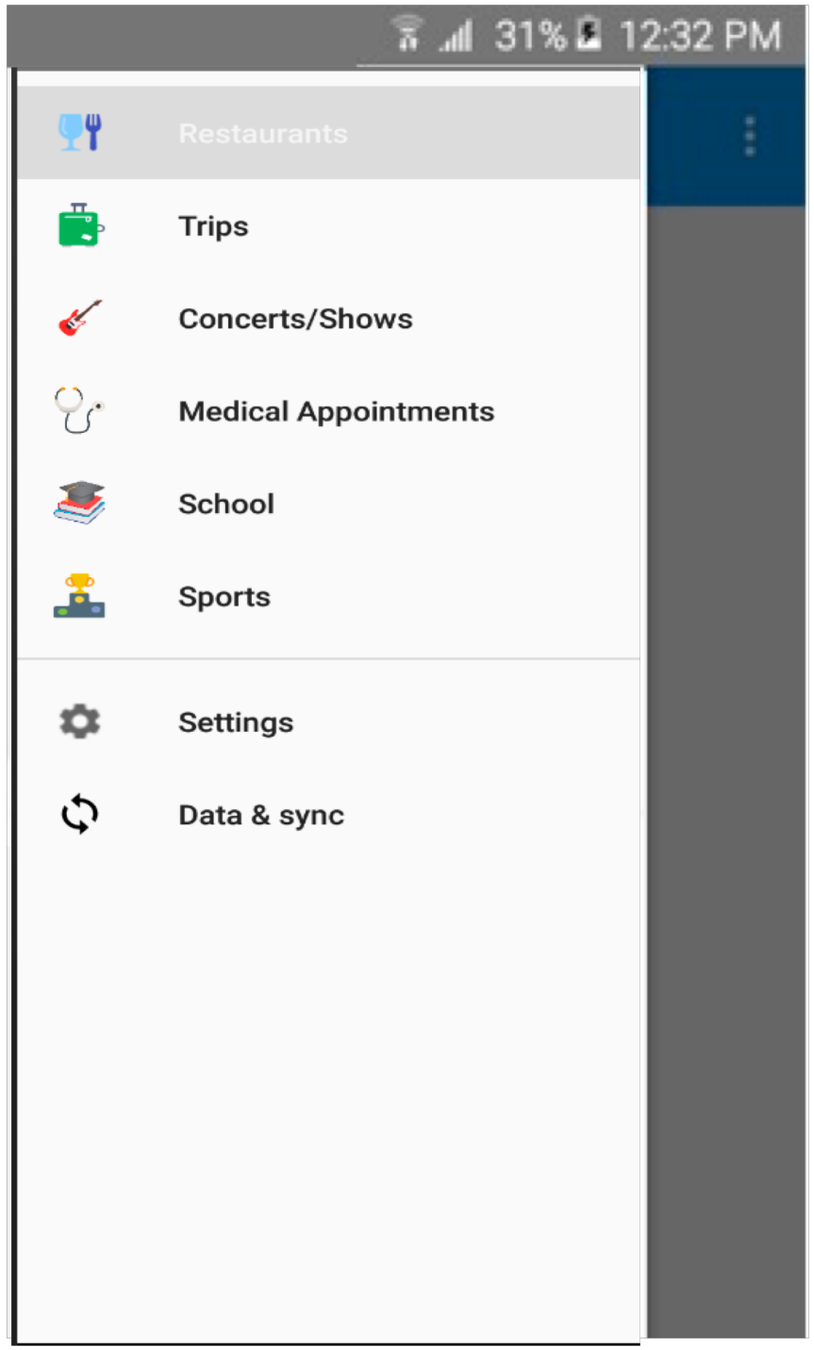


Figure: Architecture

YOUR DIGITAL SELF

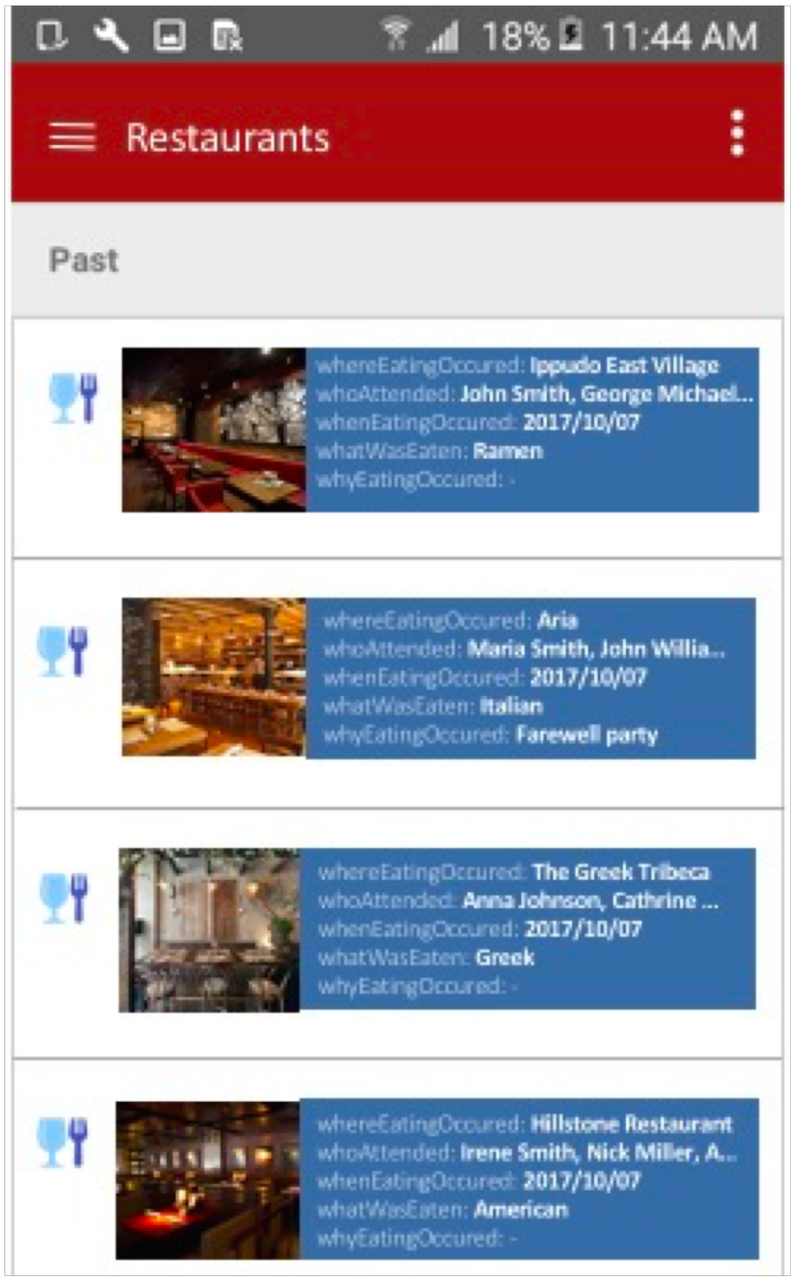


(a) List of sources

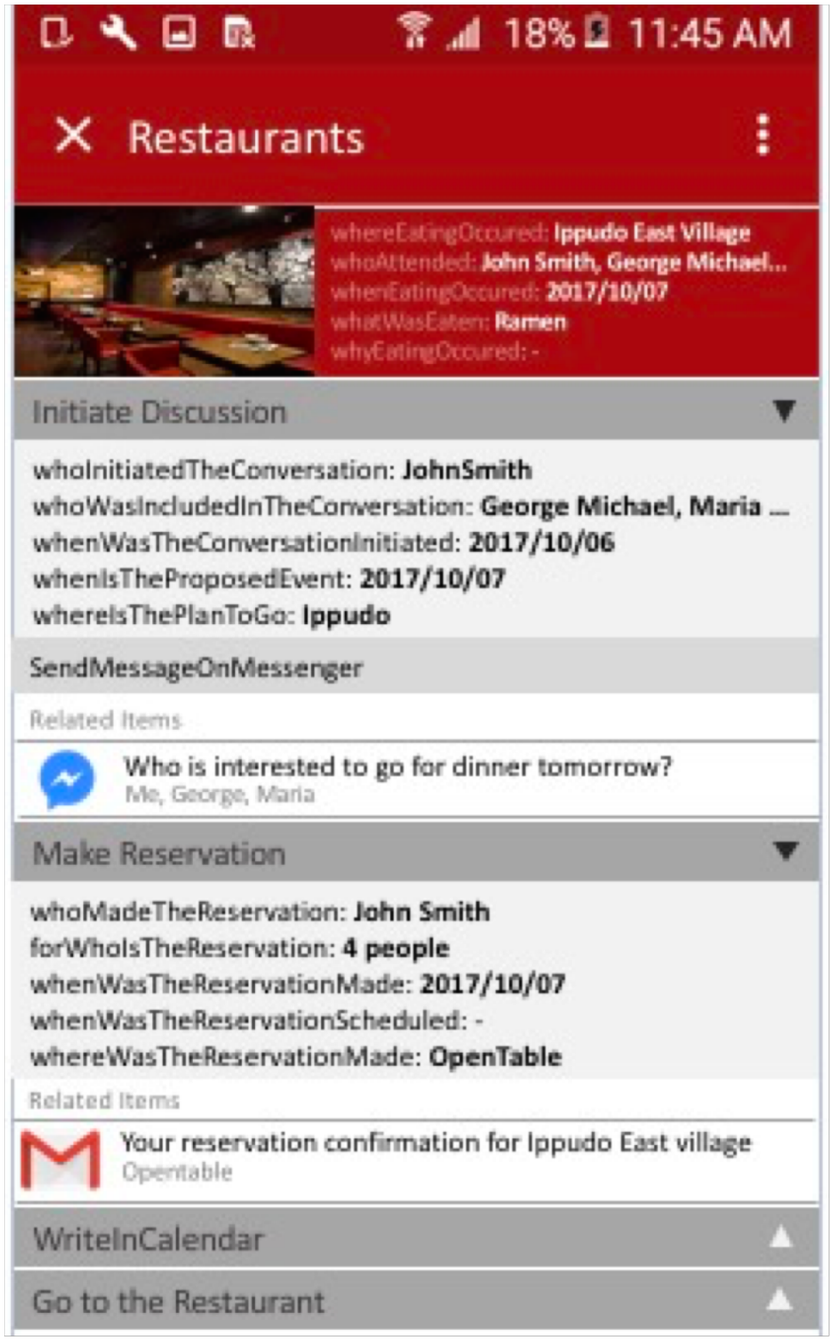


(b) List of script categories

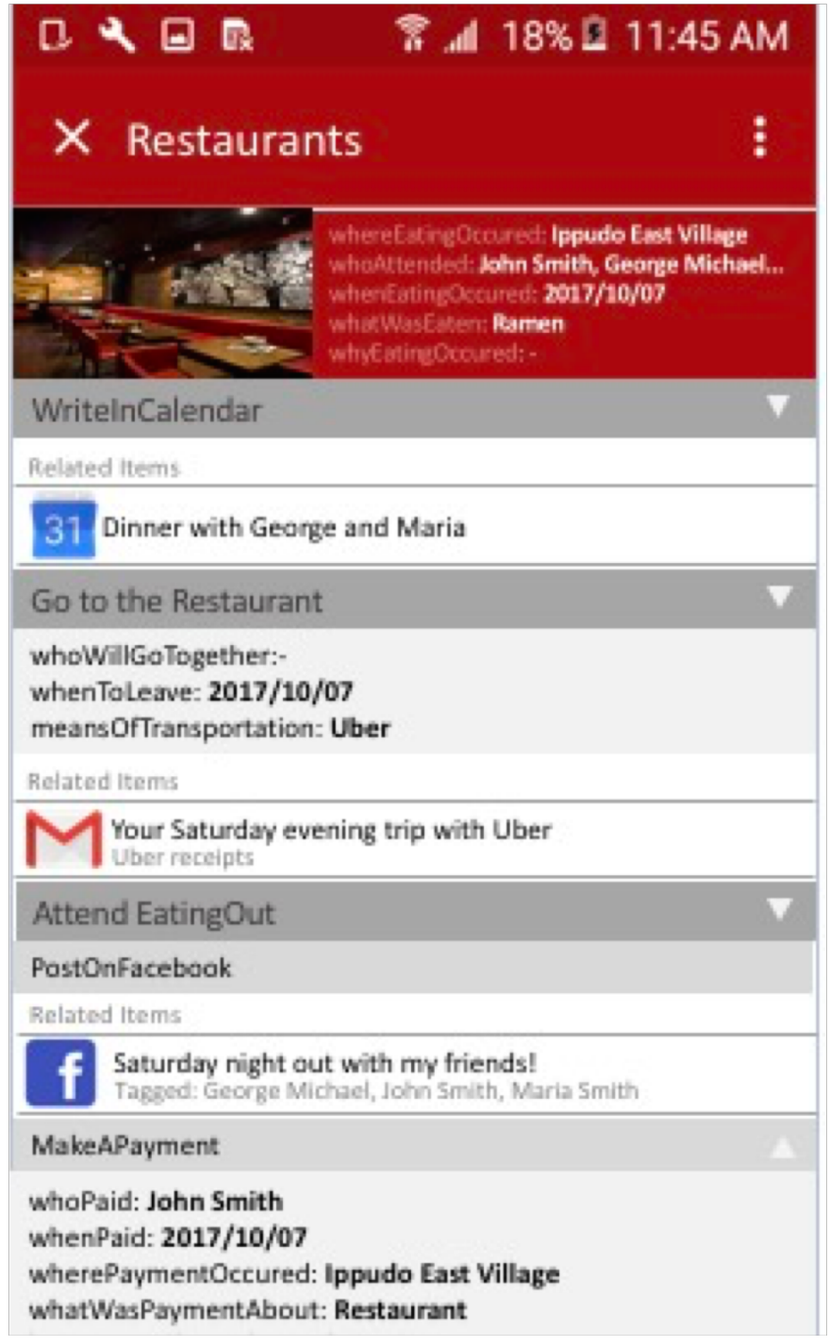
YOUR DIGITAL SELF



List of recognized restaurant outings

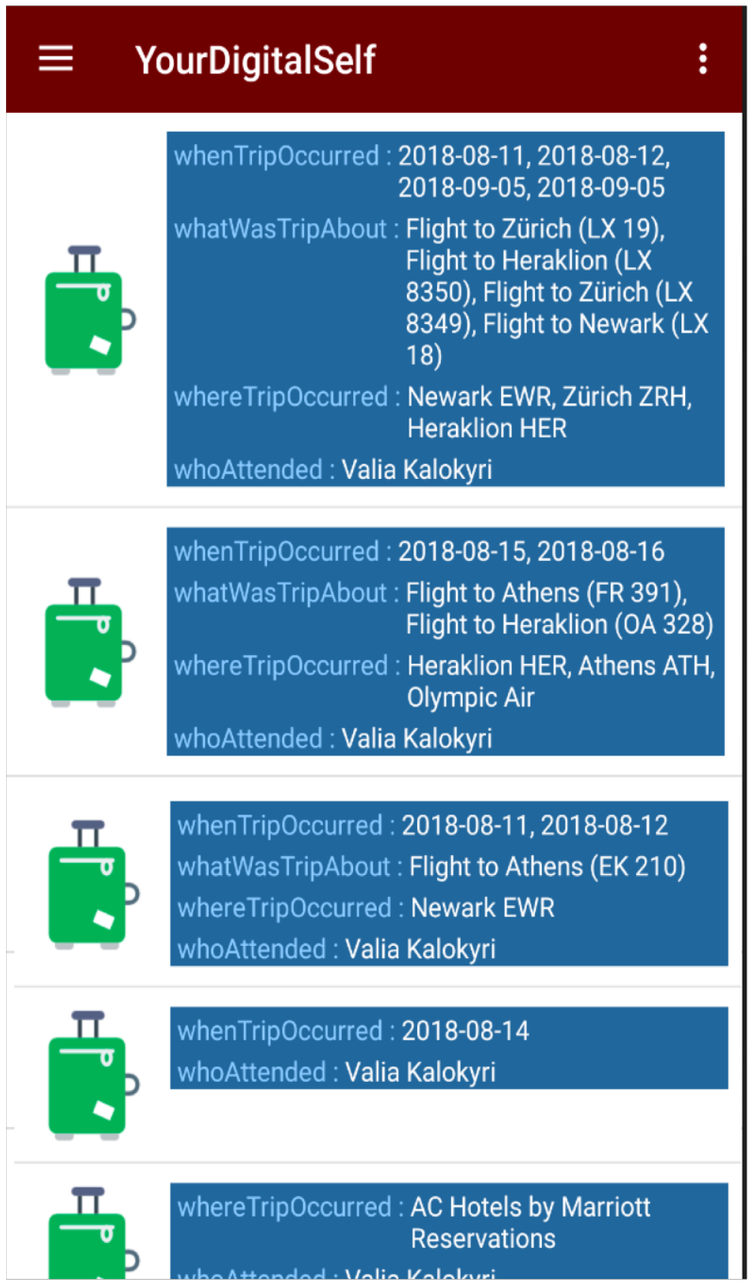


Instantiation of a restaurant outing (1/2)

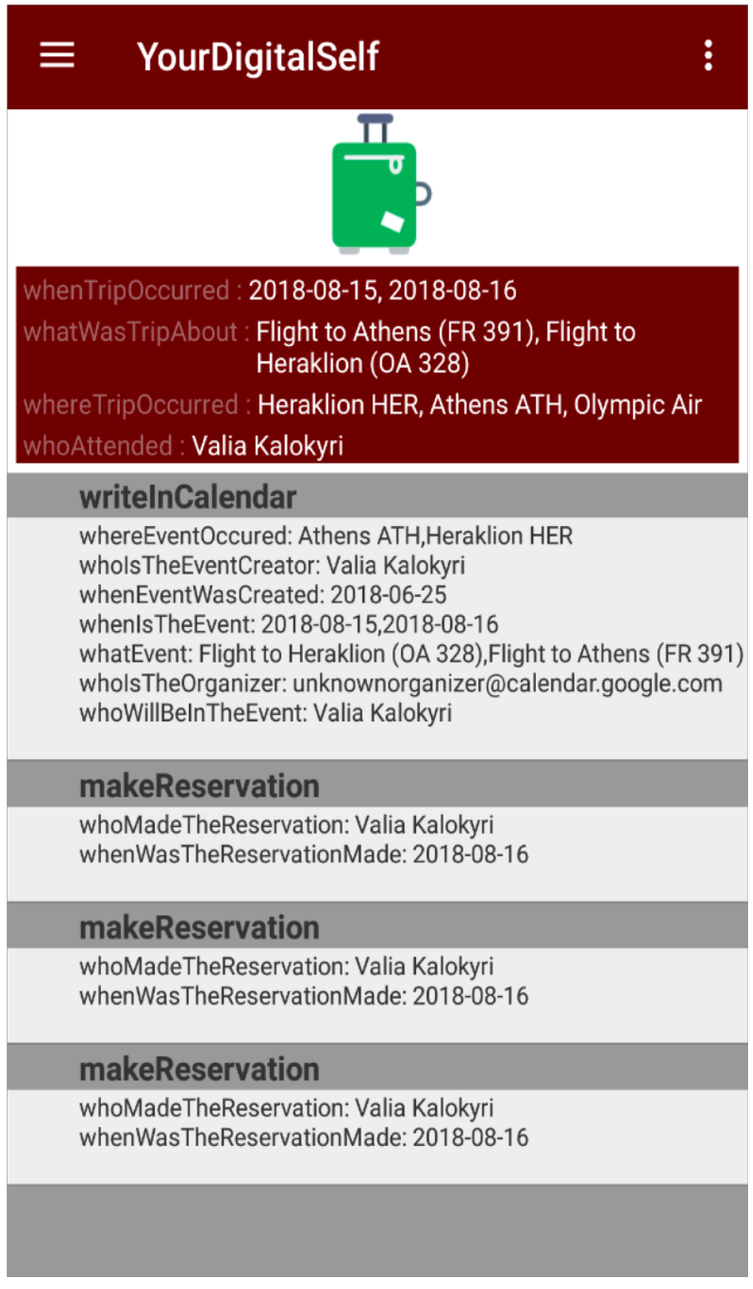


Instantiation of a restaurant outing (2/2)

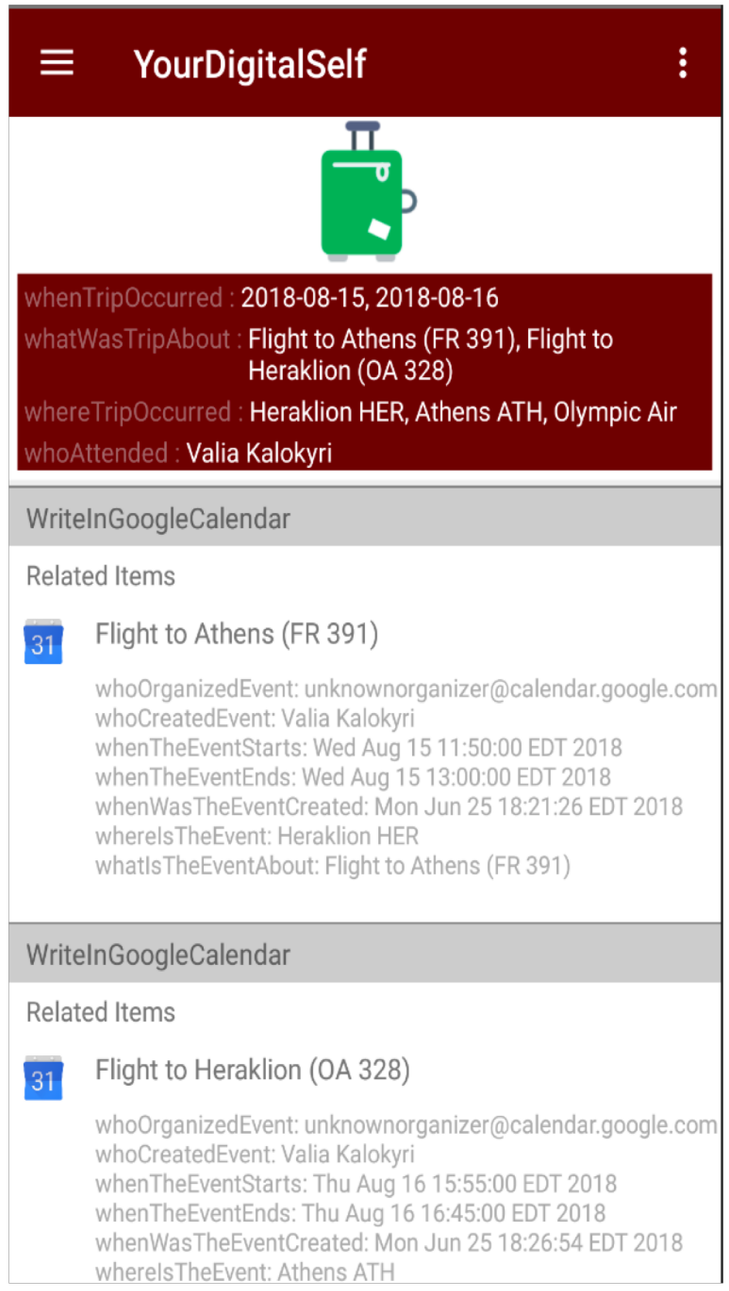
YOUR DIGITAL SELF



List of recognized trips



Instantiation of a trip instance(1/2)



Instantiation of a trip instance (2/2)

RELATED WORK

- **Personal Information Management Systems - PIMs**
  - Focus on object relationships
    - **Haystack, pStore, Semex, OntoPIM**
  - Episodic memory is much more extensive and relevant in making sense of a **collection** of documents.
- **Processes and Plans**
  - Rely on Machine Learning of patterns/process schemas from large collection of examples (sensor data).
  - In our case, a very large fraction of the plan steps in any particular instantiation of a script leave no trace ("missing actions")
- **New approach based on Information Retrieval.**

CONCLUSION

- **Integrate personal digital traces** by developing techniques to **retrieve**, **store** and **index PDTs** from various heterogeneous sources.
- **Design of a formal conceptual model and script language** for **linking** and **representing** both **PDTs** and prototypical plans (**scripts**) for common everyday events.
- **Group personal data with respect to conceptually coherent episodes** by means of **scripts**.
- **Extensible approach for implementing script instantiation** from PDTs.
  - Declarative description of scripts, evidence, clues, mappings
- **Evaluation** of our approach through a case study on **real-user** data.
- **Design of a mobile application with narrative views of users' digital memories.**

RUTGERS
School of Arts and Sciences

THANK YOU!

ANY QUESTIONS?

v.kalokyri@cs.rutgers.edu