WHYPER: Towards Automating Risk Assessment of Mobile Applications

Rahul Pandita, Xusheng Xiao, Wei Yang, William Enck, and Tao Xie♠
Department of Computer Science
North Carolina State University
♠ University of Illinois at Urbana-Champaign

Application Markets

"Application markets have played an important role in the popularity of smartphones and mobile devices."

Apple App Store

Google Play

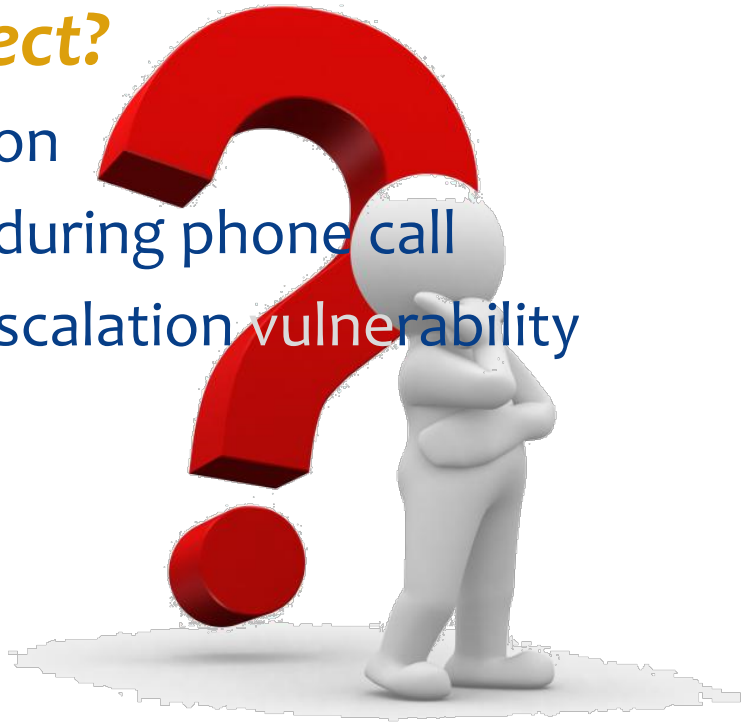
Microsoft Windows Phone

Predominant approaches towards Market Security/Privacy

- **Apple** (*Market's* Responsibility)
  - Apple performs manual inspection
- **Google** (*User's* Responsibility)
  - Users approve permissions for security/privacy
  - Bouncer (static/dynamic malware analysis)
- **Windows Phone** (Hybrid)
  - Permissions / manual inspection

Is Program Analysis sufficient?

- Previous approaches look at permissions, code, and runtime behaviors
- Caveat: ***what does the users expect?***
  - **GPS Tracker:** record and send location
  - **Phone-Call Recorder:** record audio during phone call
  - **One-Click Root**: exploit a privilege escalation vulnerability
  - Others are more subtle

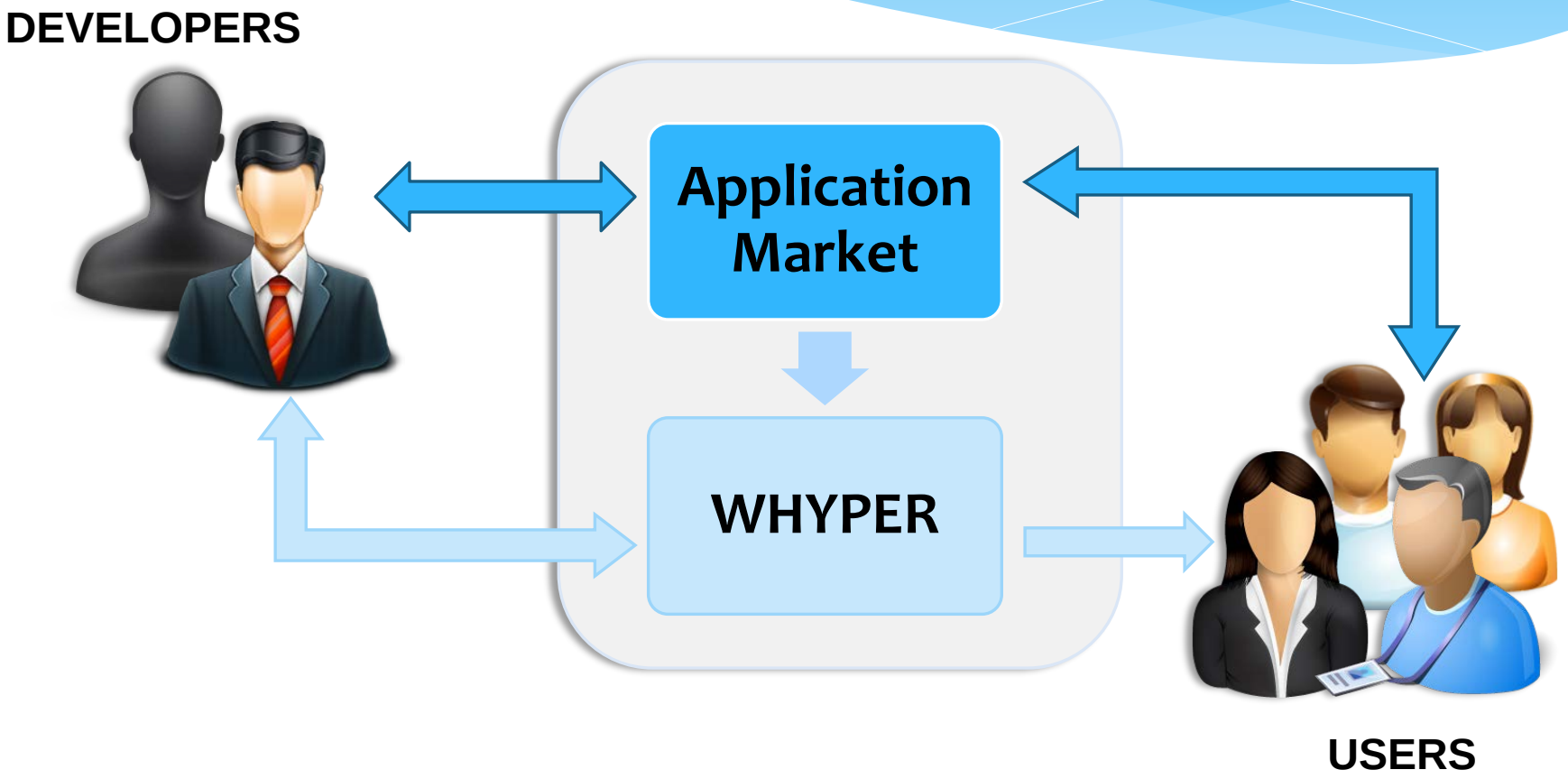


Vision

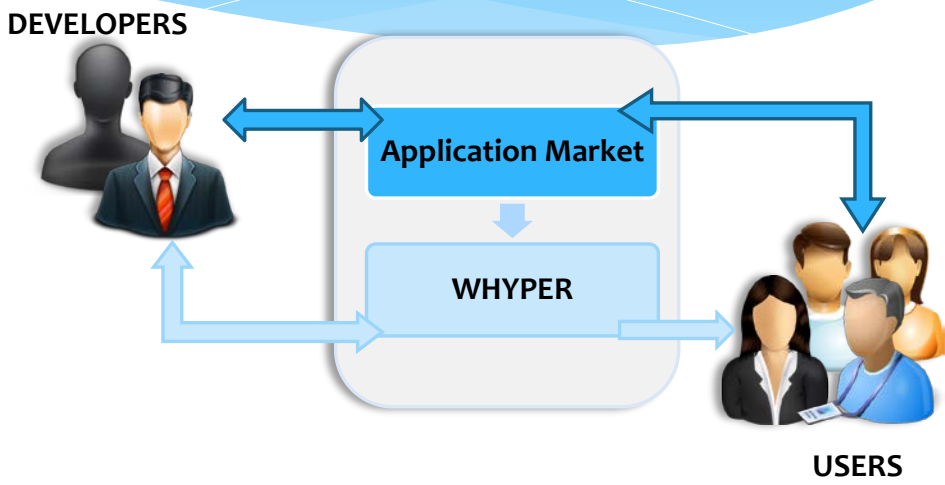
"Bridging the gap between user expectation and app behaviors"

- A first step in this direction
- Focus on permission and application descriptions
  - permissions protecting user understandable resources should be discussed
  - low-level system permissions are unlikely to be mentioned

WHYPER Overview



Use Cases



- Enhance user experience
  - *while installing Apps*
- Functionality disclosure
  - *enforce on part of developers*
- Complementing program analysis
  - *to ensure more appropriate justifications*

Solution

Simple Solution?

Keyword-based search on application descriptions

Photo Credit: Ahora estoy en via Flickr

Problems with Ctrl + F

- **Confounding effects:**
  - Certain keywords such as “contact” have a confounding meaning.
  - For instance, “... **displays user contacts**, ...” vs “... **contact me at abc@xyz.com**”.
- **Semantic Inference:**
  - Sentences often describe a sensitive operation such as reading contacts without actually referring to keyword “contact”.
    - For instance, “**share yoga exercises with your friends via email, sms**”.

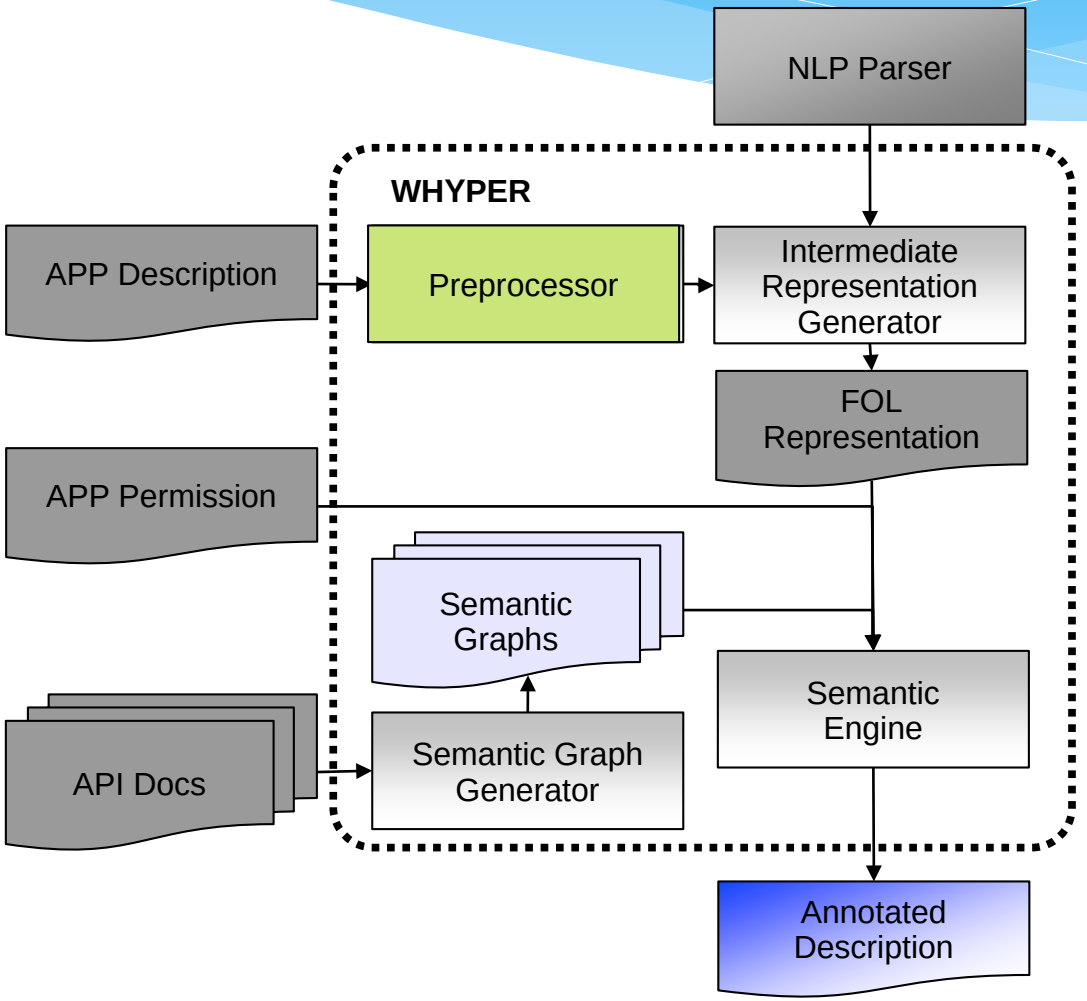
Natural Language Processing (NLP)

- NLP techniques help computers understand NL artifacts
- NLP is still difficult
- NLP on domain specific sentences with specific styles is feasible

NLP Preliminaries

- **Parts Of Speech (POS) Tagging**
  - E.g., noun, verb, prepositions…
- **Phrase and Clause Parsing**
  - E.g., noun phrases (basketball players) and verb phrases (make sure)…
- **Stanford-Typed Dependencies**
  - E.g., subject, object , adverbial modifiers…
- **Named Entity Recognition**
  - E.g., 'Pandora Internet Radio' is a name, '$5' refers to a currency amount…

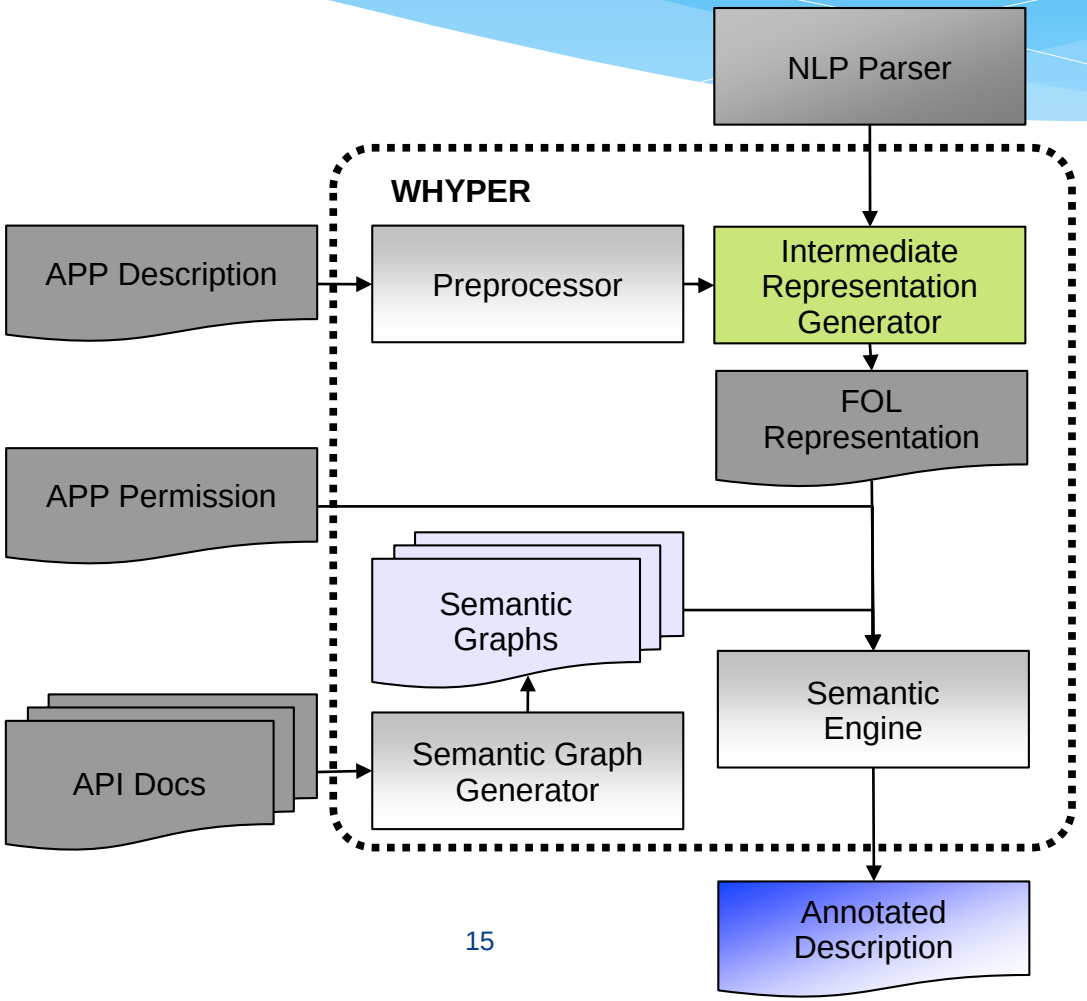
WHYPER Framework



Preprocessor

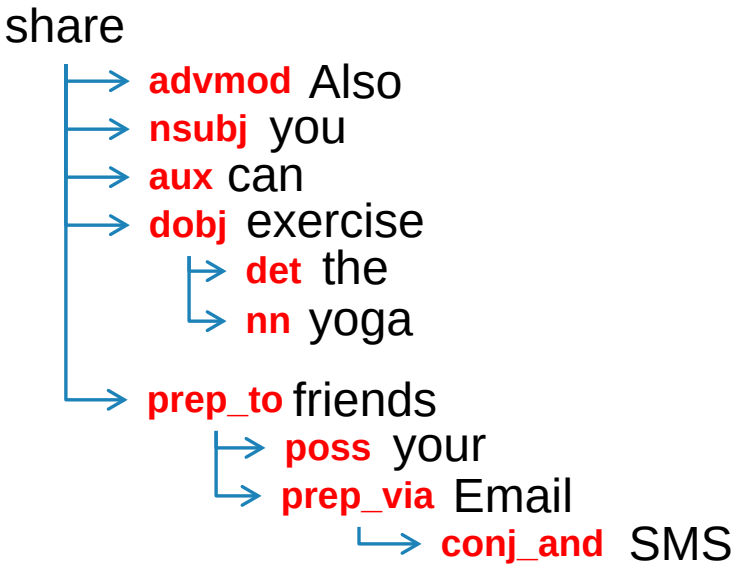
- **Period Handling**
  - Decimals, ellipsis, shorthand notations (**`Mr., Dr.`**)
- **Sentence Boundaries**
  - Tabs, bullet points, delimiters (:)
  - Symbols (*,-) and enumeration sentence
- **Named Entity Handling**
  - *E.g.*, "**`Pandora internet radio`**",
- **Abbreviation Handling**
  - *E.g.*, "**`Instant Message (IM)`**"

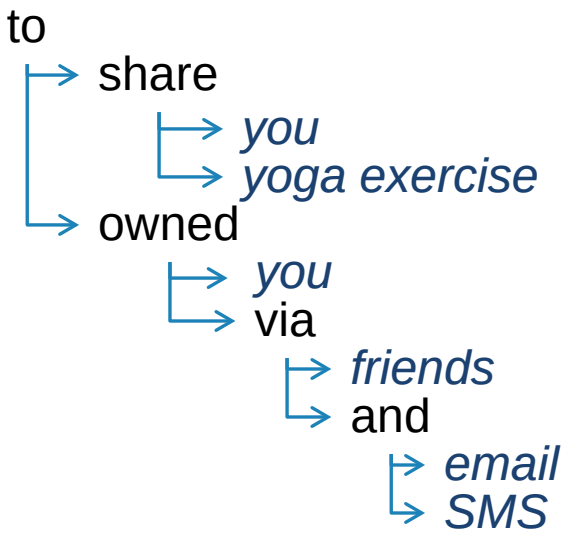
WHYPER Framework



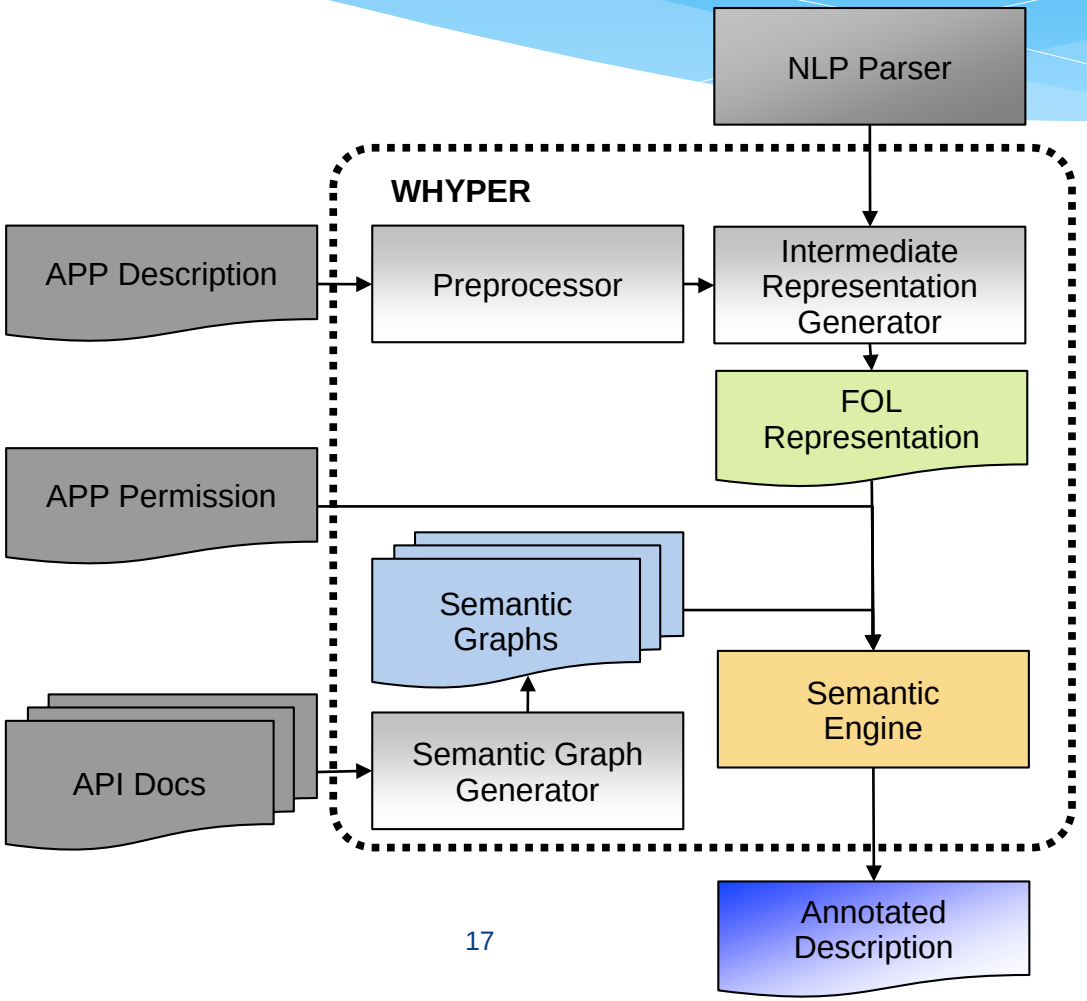
"Also you can share the yoga exercise to your friends via Email and SMS."

Also	you	can	share	the	yoga	exercise	to	your	friends	via	Email	and	SMS
RB	**PRP**	**MD**	**VB**	**DT**	**NN**	**NN**	**PRP**	**NNS**		**NNP**		**NNP**	

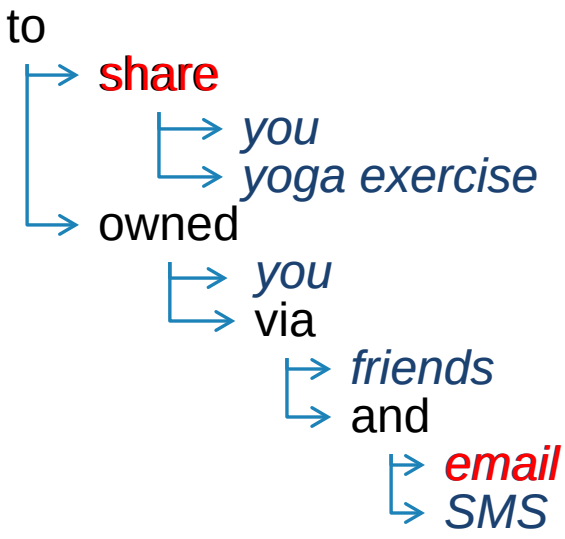




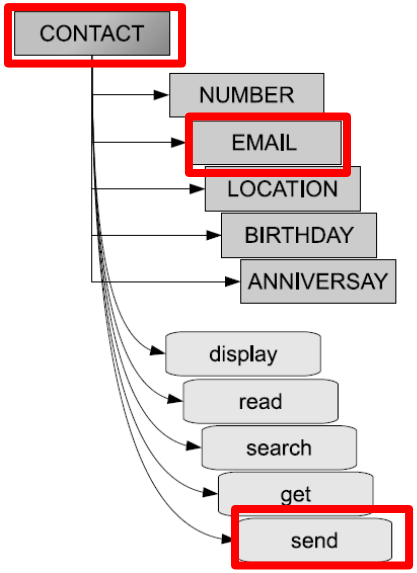
WHYPER Framework



"Also you can share the yoga exercise to your friends via Email and SMS."



WordNet Similarity

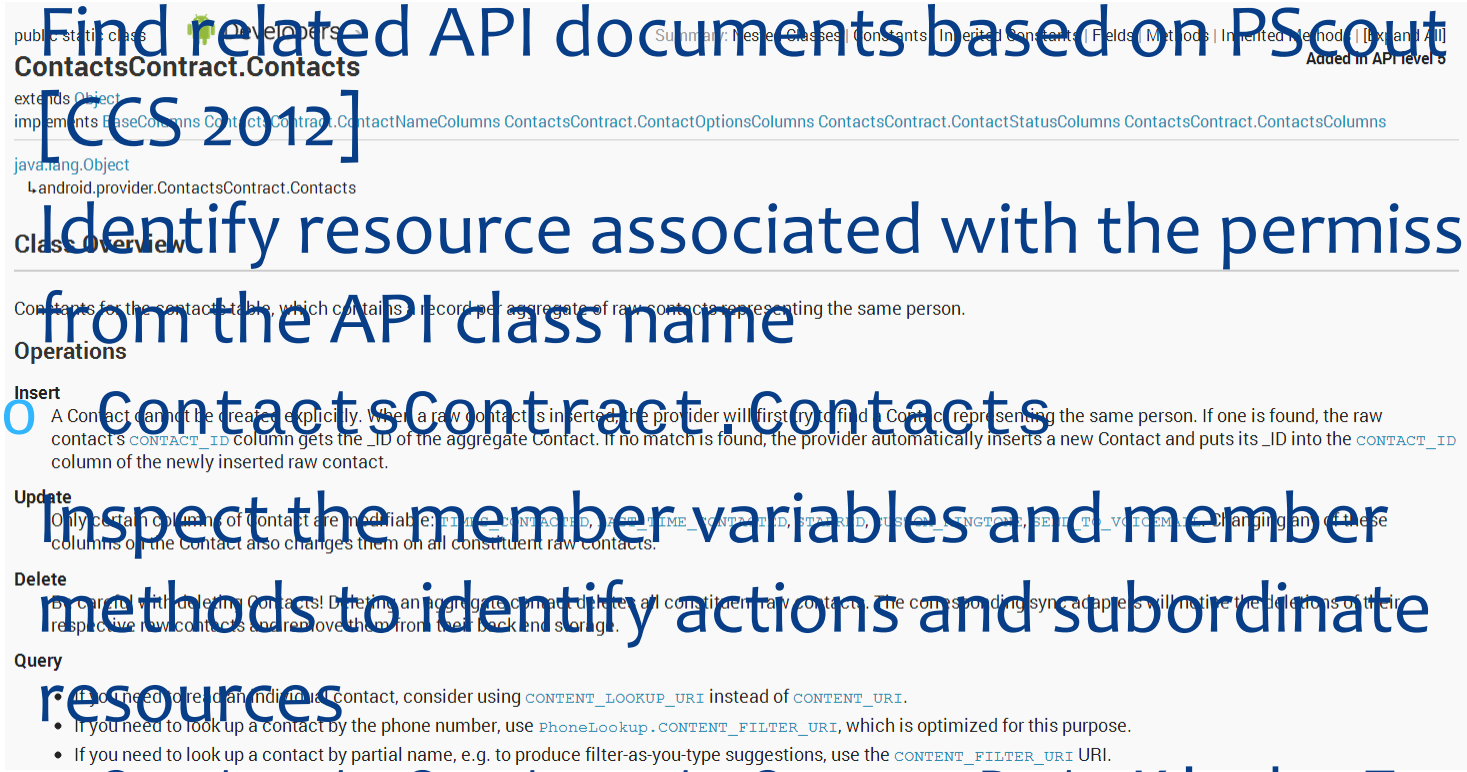


Semantic-Graph Generator

- **WHY**
  - *to perform deep semantic analysis*
- **HOW**
  - *infer graphs from API documents*

Semantic-Graph Generator

- **Systematic approach to infer graphs**
  - Find related API documents based on PScout [CCS 2012]
  - Identify resource associated with the permissions from the API class name
    - ContactsContract.Contacts
  - Inspect the member variables and member methods to identify actions and subordinate resources
    - ContactsContract.CommonDataKinds.Email



Evaluation

- **Subjects**
  - Permissions:
    - READ_CONTACTS
    - READ_CALENDAR
    - RECORD_AUDIO
  - 581/600* application descriptions (only English descriptions)
  - **9,953** sentences
- **Research Questions (RQs)**
  - **RQ1**: What are the precision, recall and F-Score of WHYPER in identifying permission sentences?
  - **RQ2**: How effective WHYPER is in identifying permission sentences, compared to keyword-based searching ?

Statistics of Subject Applications

Permissions	#N	#S	S_p
READ_CONTACTS	190	3,379	235
READ_CALENDAR	191	2,752	283
RECORD_AUDIO	200	3,822	245
TOTAL	**581**	**9,953**	**763**

Classification

- **True Positive (TP):**
  - WHYPER(*sentence*) && Manual(*sentence*)
- **False Positive (FP):**
  - WHYPER(*sentence*) && ! Manual(*sentence*)
- **True Negative (TN):**
  - ! WHYPER(*sentence*) && ! Manual(*sentence*)
- **False Negative (FN):**
  - ! WHYPER(*sentence*) && Manual(*sentence*)

RQ1 Results: Effectiveness of WHYPER

Permission	S_I	TP	FP	FN	TN	Prec.	Recall	F-Score	Acc
READ_CONTACTS	204	186	18	49	2,930	91.2	79.2	84.8	97.9
READ_CALENDAR	288	241	47	42	2,422	83.7	85.2	84.5	96.8
RECORD_AUDIO	259	195	64	50	3,470	75.3	79.6	77.4	97.0
TOTAL	**751**	**622**	**129**	**141**	**9,061**	**82.8**	**81.5**	**82.2**	**97.3**

- **Low FPs and FNs**
  - out of *9,061* sentences, only *129* are flagged as **FP**s
  - among *581* applications, *109* applications (18.8%) contain at least one **FP**
  - among *581* applications, *86* applications (14.8%) contain at least one **FN**

RQ2 Results: Comparison to Keyword-based Search

Permission	Keywords
READ_CONTACTS	contact, data, number, name, email
READ_CALENDAR	calendar, event, date, month, day, year
RECORD_AUDIO	record, audio, voice, capture, microphone

RQ2 Results: Comparison to Keyword-based Search

Permission	Delta Precision	Delta Recall	Delta F-score	Delta Accuracy
READ_CONTACTS	50.4	1.3	31.2	7.3
READ_CALENDAR	39.3	1.5	26.4	9.2
RECORD_AUDIO	36.9	-6.6	24.3	6.8
WHYPER Improvement	**41.6**	**-1.2**	**27.2**	**7.7**

Result Analysis (False Positives)

- **Incorrect parsing**
  - *"MyLink Advanced provides full synchronization of all Microsoft Outlook emails (inbox, sent, outbox and drafts), contacts, calendar, tasks and notes with all Android phones via USB"*
- **Synonym analysis**
  - *"You can now turn recordings into ringtones."*

Result Analysis (False Negatives)

- **Incorrect parsing**
  - Incorrect identification of sentence boundaries and limitations of underlying NLP infrastructure
- **Limitations of Semantic Graphs**
  - **Manual Augmentation**
    - microphone-*blow into* and call-*record*
    - *significant improvement of Delta Recalls: -6.6% to 0.6%*
  - Automatic mining from user comments and forums

Discussions

- **Generalization to other permissions**
  - user-understandable permissions: calls, SMS
  - location and phone identifiers
  - internet

Conclusion

- We propose the use of NLP techniques to ***help bridge the semantic gap*** between what mobile applications do and what users expect them to do.
- Our evaluation demonstrates an ***improvement*** over a simple keyword-based searching.

Thank You