Toward the Automated Localization of Buggy Mobile App UIs from Bug Descriptions

Antu Saha



Yang Song



Junayed Mahmud



Ying Zhou



Kevin Moran



Oscar Chaparro



WILLIAM & MARY
CHARTERED 1693

UCF
UNIVERSITY OF CENTRAL FLORIDA

GEORGE MASON UNIVERSITY

Problem Motivation



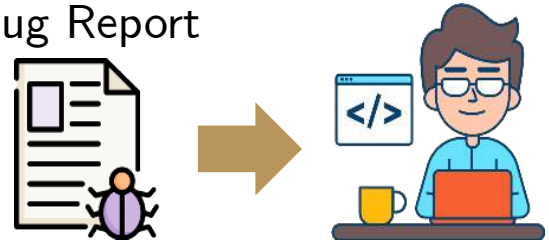
Problem Motivation

Bug Report



Problem Motivation

Bug Report



Problem Motivation

Bug Report





Android App



Problem Motivation

Bug Report





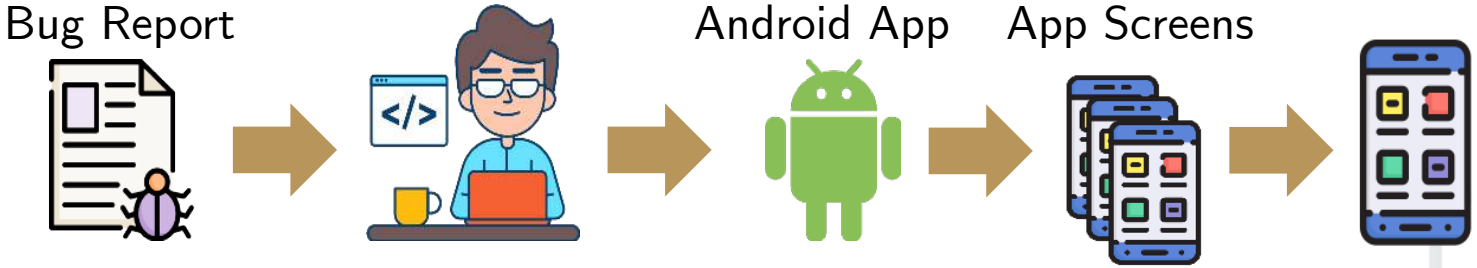
Android App



App Screens



Problem Motivation



Problem Motivation



Problem Motivation



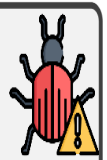
Example



Example

WifiAnalyzer App

Title: Can no longer enter text in SSID Filter TextView



Description: Cannot enter any text in the SSID Filter field.

Steps:

1. Click on Filter icon.
2. Click/tap on SSID Filter text field.
3. Keyboard does not pop up.

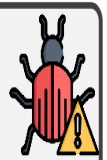
Expected Behaviour:

Should display keyboard and allow you to enter SSID filter text.

Example

WifiAnalyzer App

Title: Can no longer enter text in SSID Filter TextView



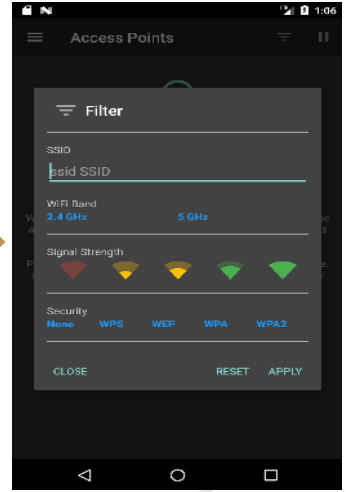
Description: Cannot enter any text in the SSID Filter field.

Steps:

1. Click on Filter icon.
2. Click/tap on SSID Filter text field.
3. Keyboard does not pop up.

Expected Behaviour:

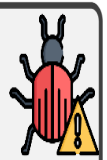
Should display keyboard and allow you to enter SSID filter text.



Example

WifiAnalyzer App

Title: Can no longer enter text in SSID Filter TextView



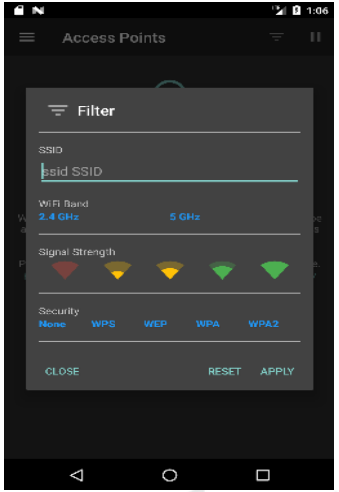
Description: Cannot enter any text in the SSID Filter field.

Steps:

1. Click on Filter icon.
2. Click/tap on SSID Filter text field.
3. Keyboard does not pop up.

Expected Behaviour:

Should display keyboard and allow you to enter SSID filter text.

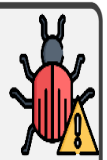


Buggy UI Screen

Example

WifiAnalyzer App

Title: Can no longer enter text in SSID Filter TextView



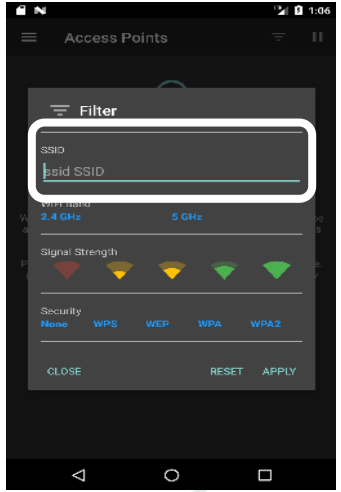
Description: Cannot enter any text in the SSID Filter field.

Steps:

1. Click on Filter icon.
2. Click/tap on SSID Filter text field.
3. Keyboard does not pop up.

Expected Behaviour:

Should display keyboard and allow you to enter SSID filter text.

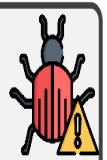


Buggy UI Screen

Example

WifiAnalyzer App

Title: Can no longer enter text in SSID Filter TextView



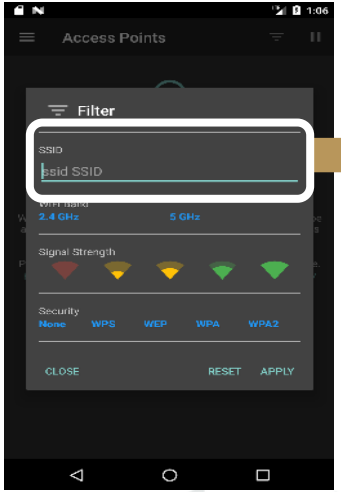
Description: Cannot enter any text in the SSID Filter field.

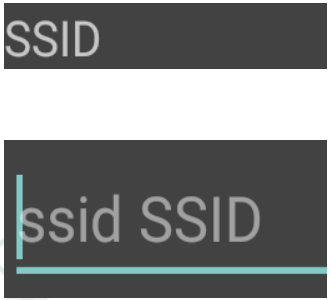
Steps:

1. Click on Filter icon.
2. Click/tap on SSID Filter text field.
3. Keyboard does not pop up.

Expected Behaviour:

Should display keyboard and allow you to enter SSID filter text.



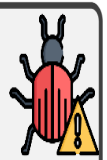


Buggy UI Screen

Example

WifiAnalyzer App

Title: Can no longer enter text in SSID Filter TextView



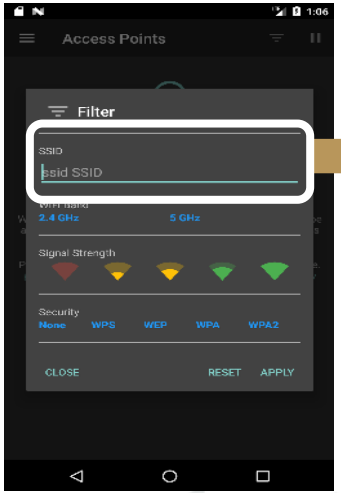
Description: Cannot enter any text in the SSID Filter field.

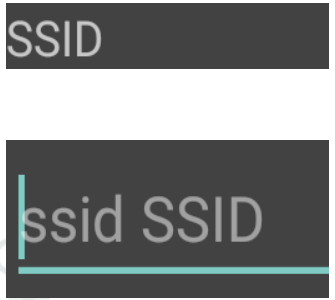
Steps:

1. Click on Filter icon.
2. Click/tap on SSID Filter text field.
3. Keyboard does not pop up.

Expected Behaviour:

Should display keyboard and allow you to enter SSID filter text.





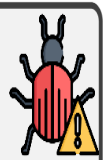
Buggy UI Screen

Buggy UI Component

Example

WifiAnalyzer App

Title: Can no longer enter text in SSID Filter TextView



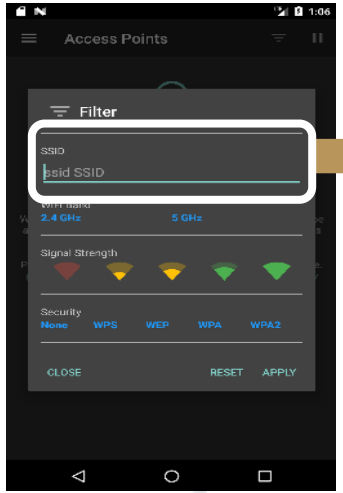
Description: Cannot enter any text in the SSID Filter field.

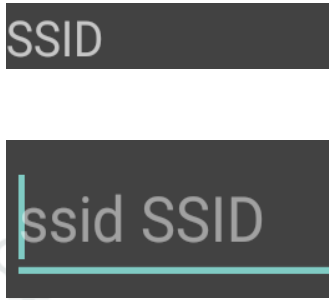
Steps:

1. Click on Filter icon.
2. Click/tap on SSID Filter text field.
3. Keyboard does not pop up.

Expected Behaviour:

Should display keyboard and allow you to enter SSID filter text.





Observed Behavior (OB)
OR
Bug Descriptions

Buggy UI Screen

Buggy UI Component

Problem Motivation



Problem Motivation

Why is manual localization of Buggy UI difficult?



Problem Motivation

Why is manual localization of Buggy UI difficult?

Low-Quality Bug Report

Bug Report





Problem Motivation

Why is manual localization of Buggy UI difficult?

Low-Quality Bug Report

Bug Report



Large Number of App Screens



Problem Motivation

Why is manual localization of Buggy UI difficult?

Low-Quality Bug Report

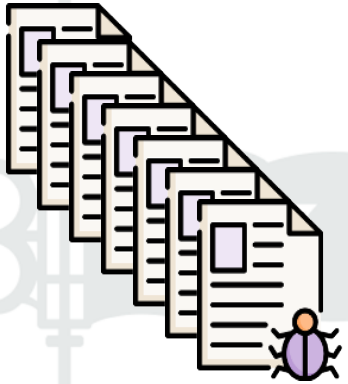
Bug Report



Large Number of App Screens



Volume of Reported Bugs



Automated Buggy UI Localization

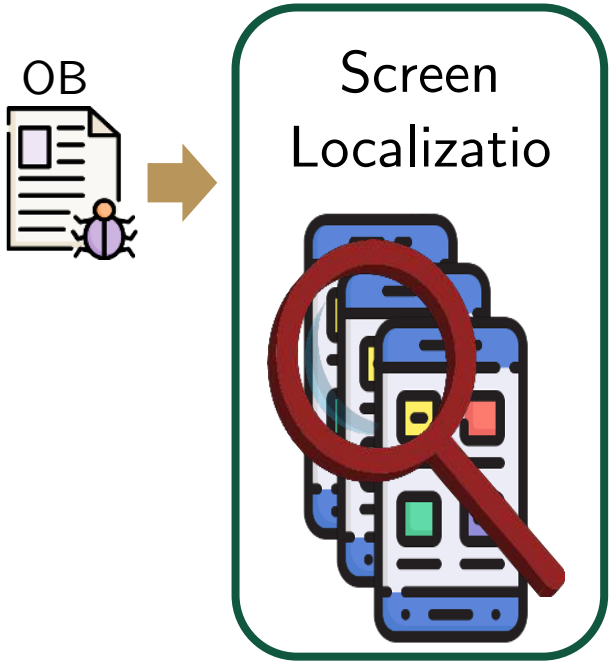


Automated Buggy UI Localization

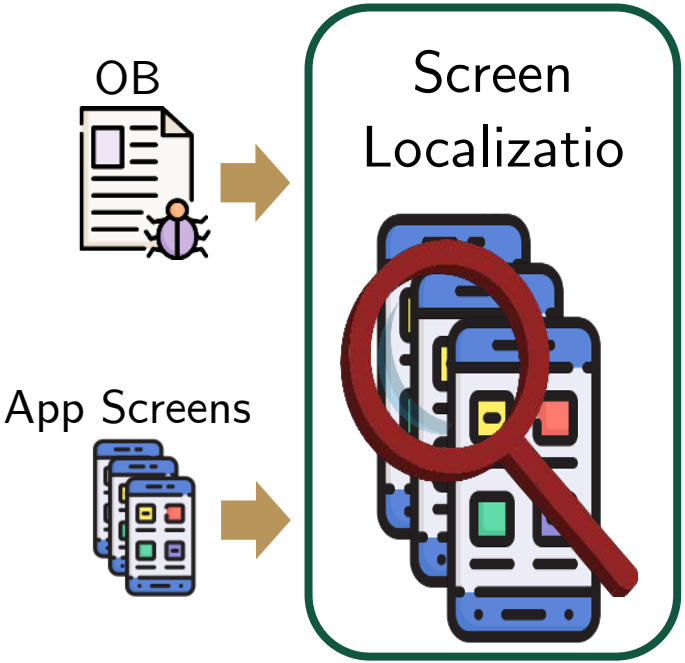
Screen Localizatio



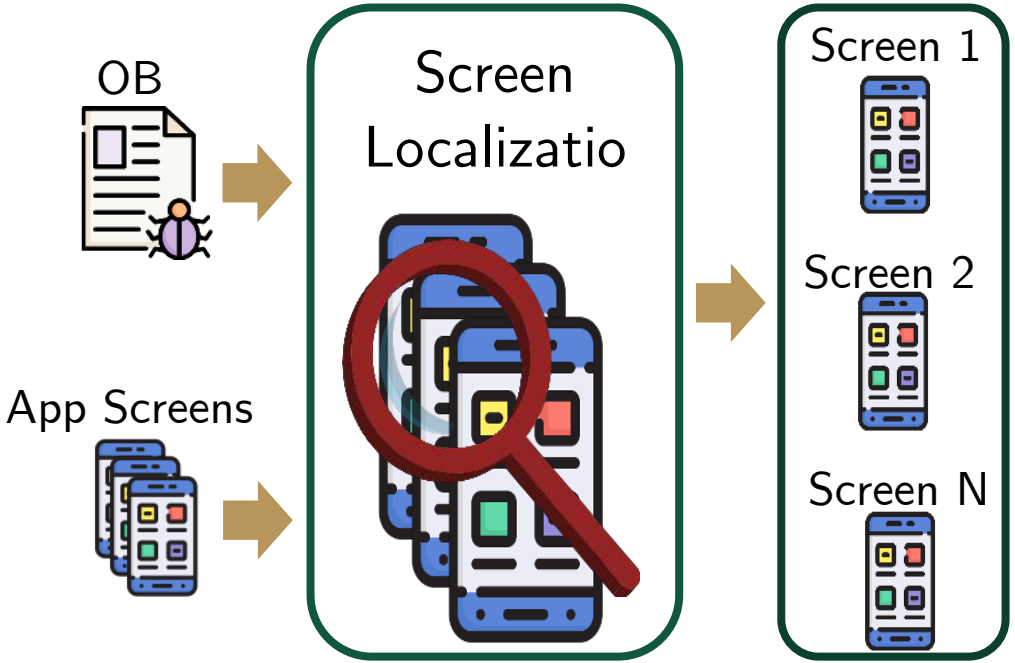
Automated Buggy UI Localization



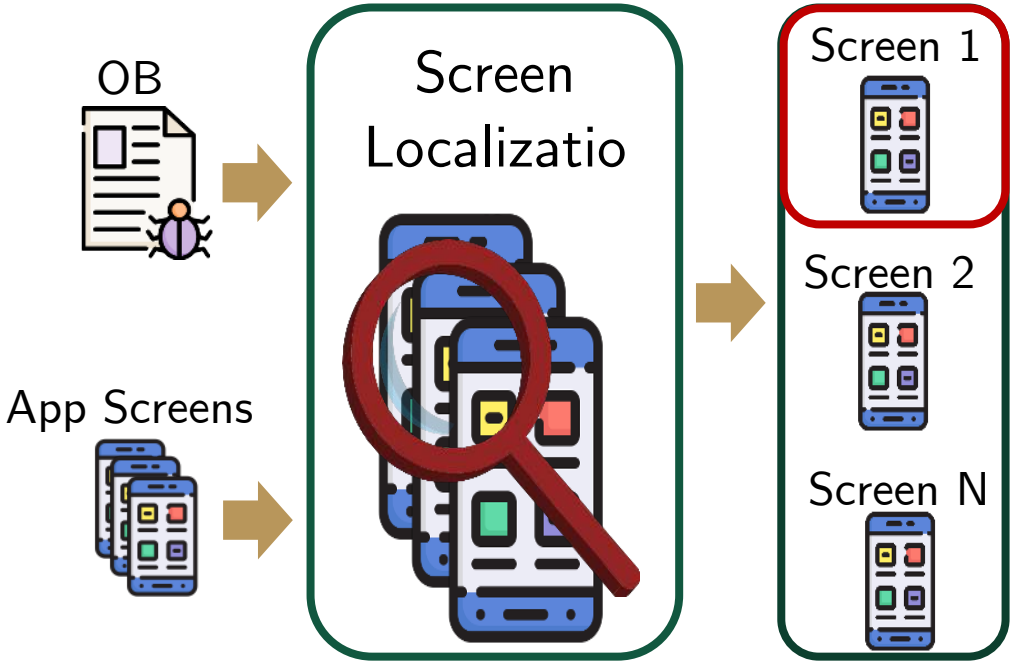
Automated Buggy UI Localization



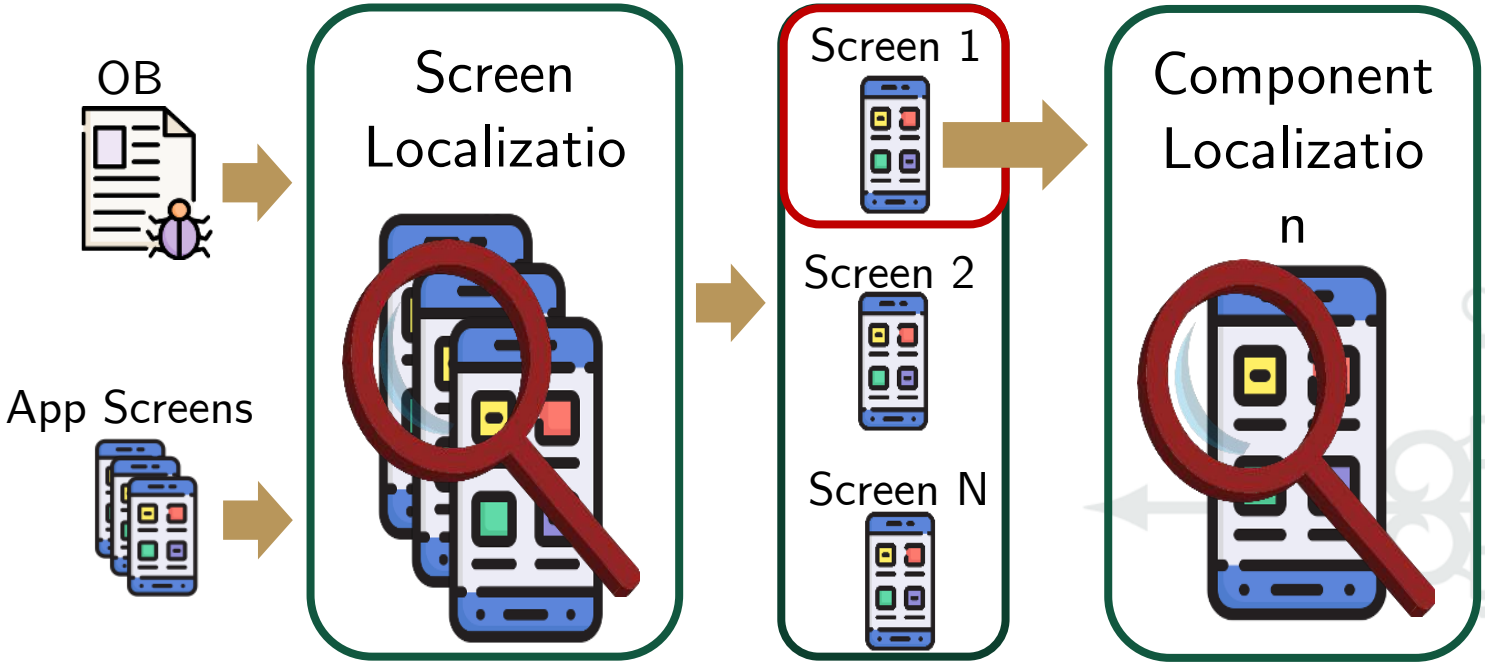
Automated Buggy UI Localization



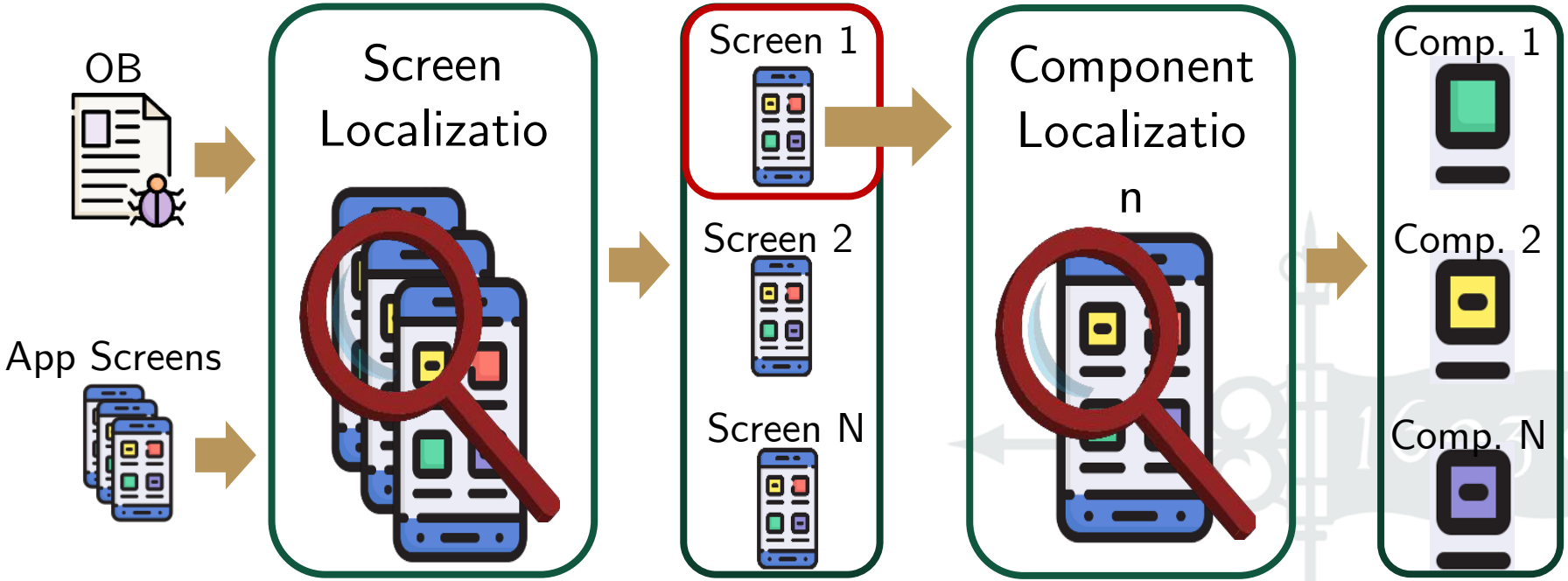
Automated Buggy UI Localization



Automated Buggy UI Localization



Automated Buggy UI Localization

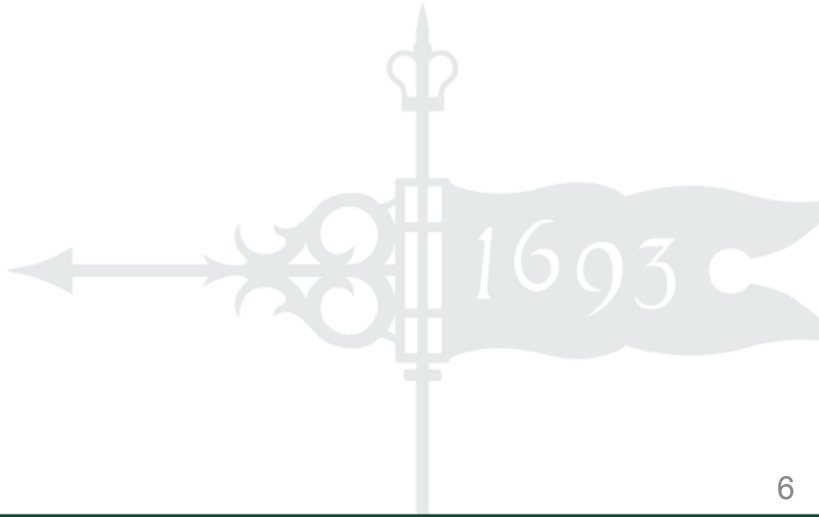


How?



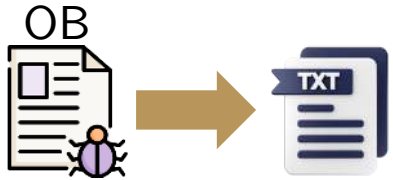
How?

Information Sources for Buggy UI Localization



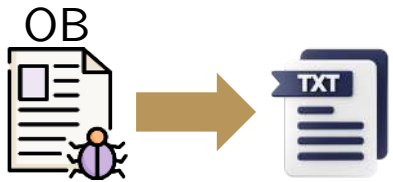
How?

Information Sources for Buggy UI Localization



How?

Information Sources for Buggy UI Localization

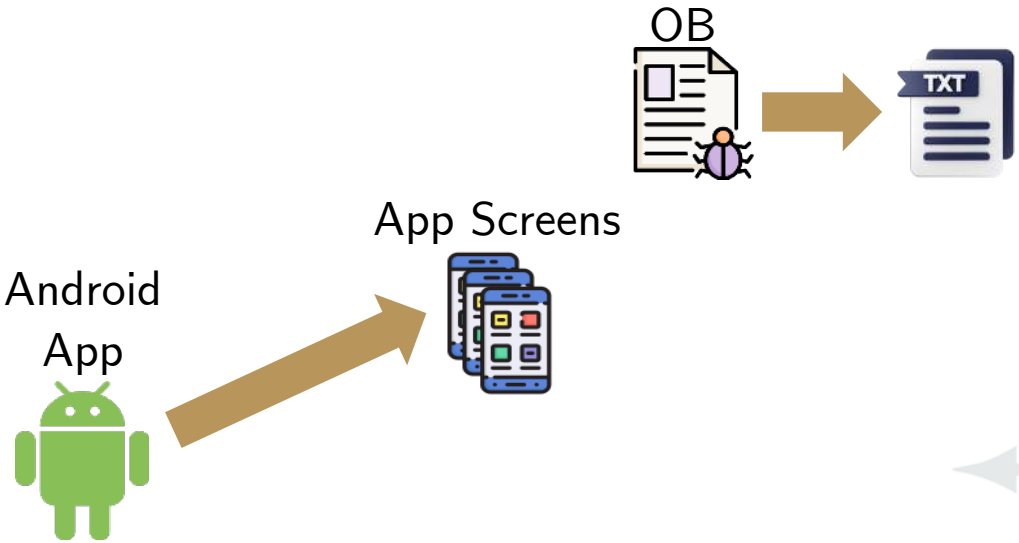


Android App



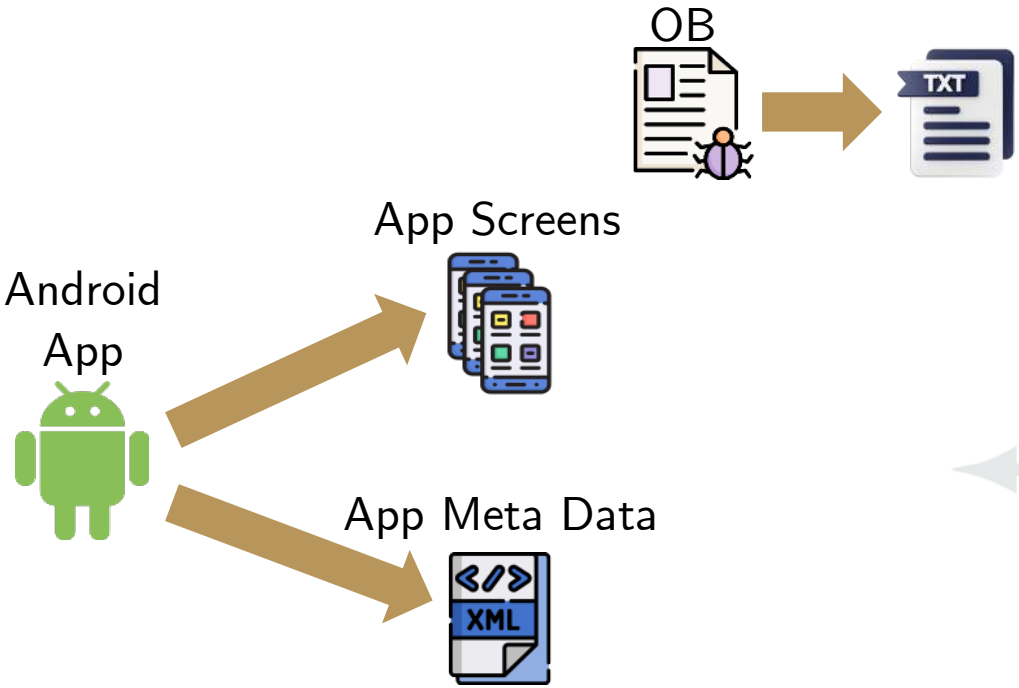
How?

Information Sources for Buggy UI Localization



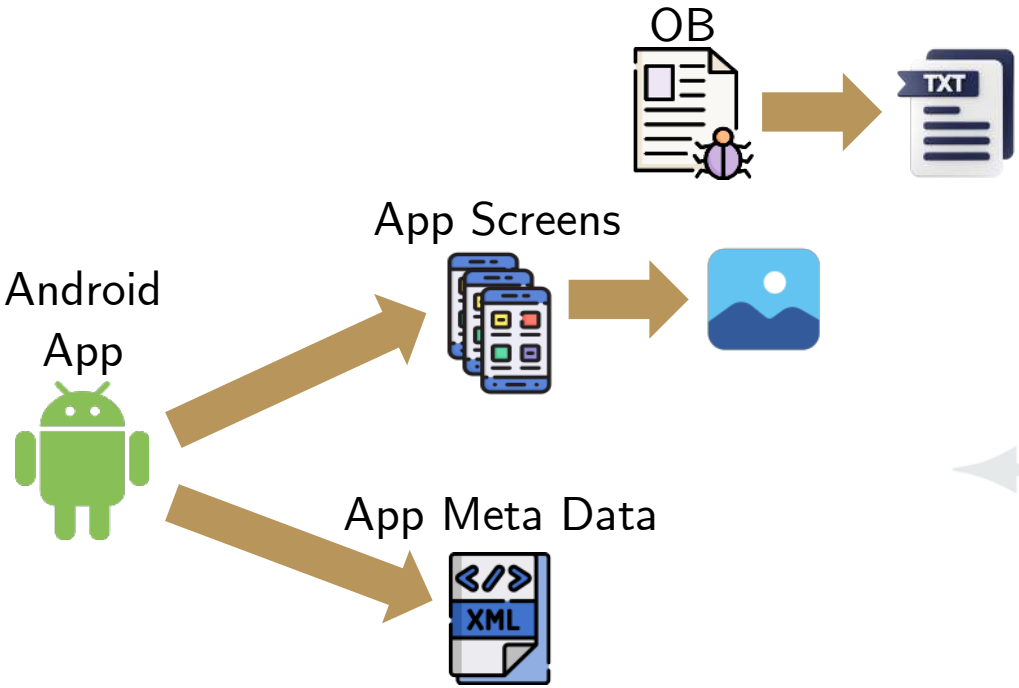
How?

Information Sources for Buggy UI Localization



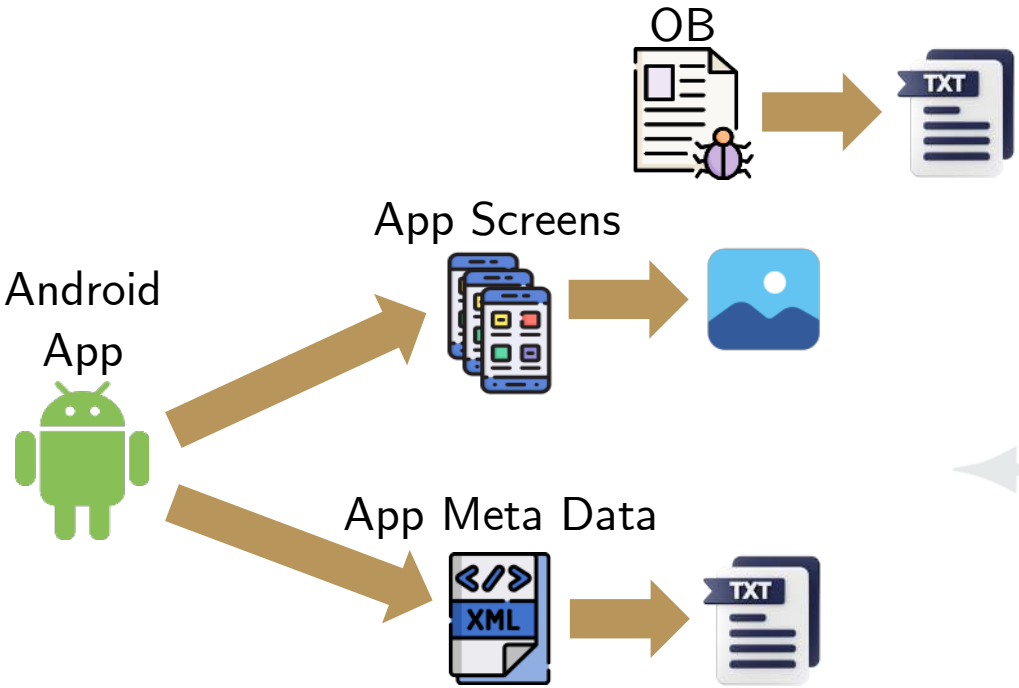
How?

Information Sources for Buggy UI Localization



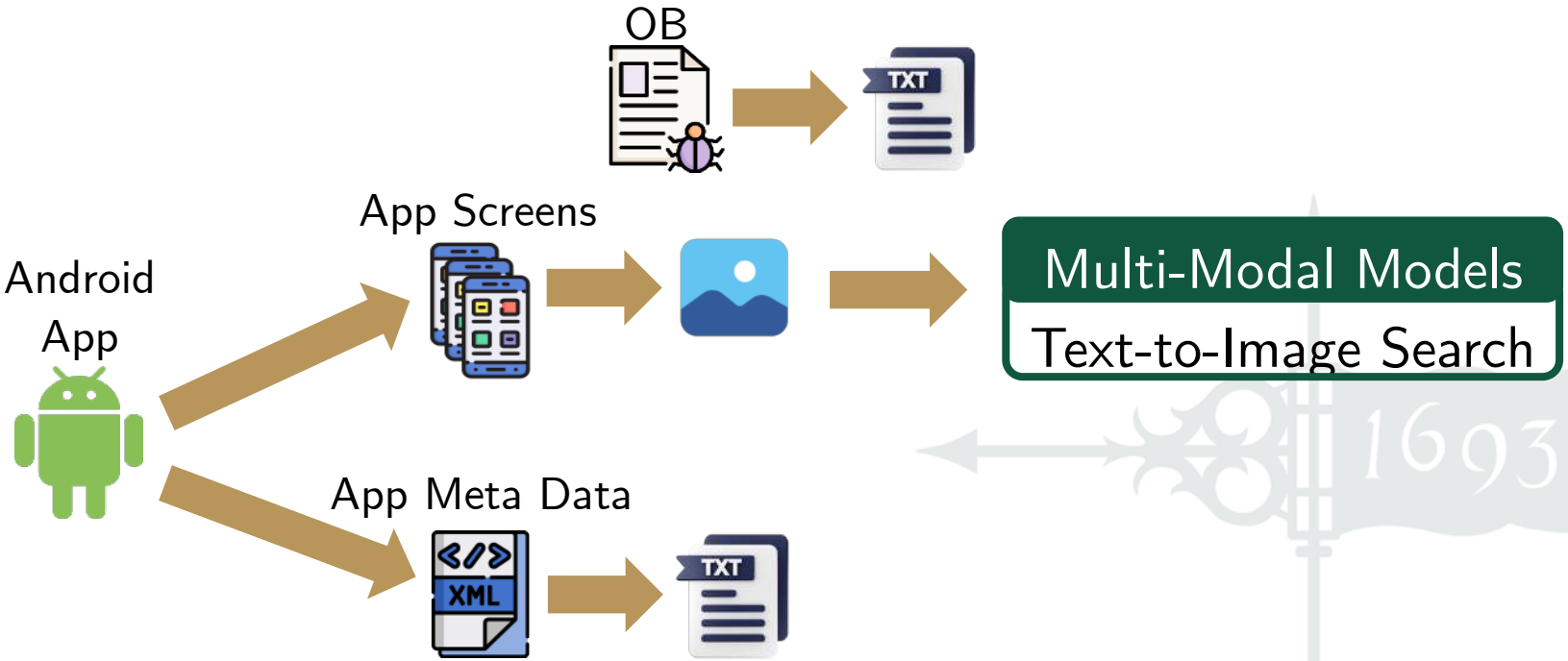
How?

Information Sources for Buggy UI Localization



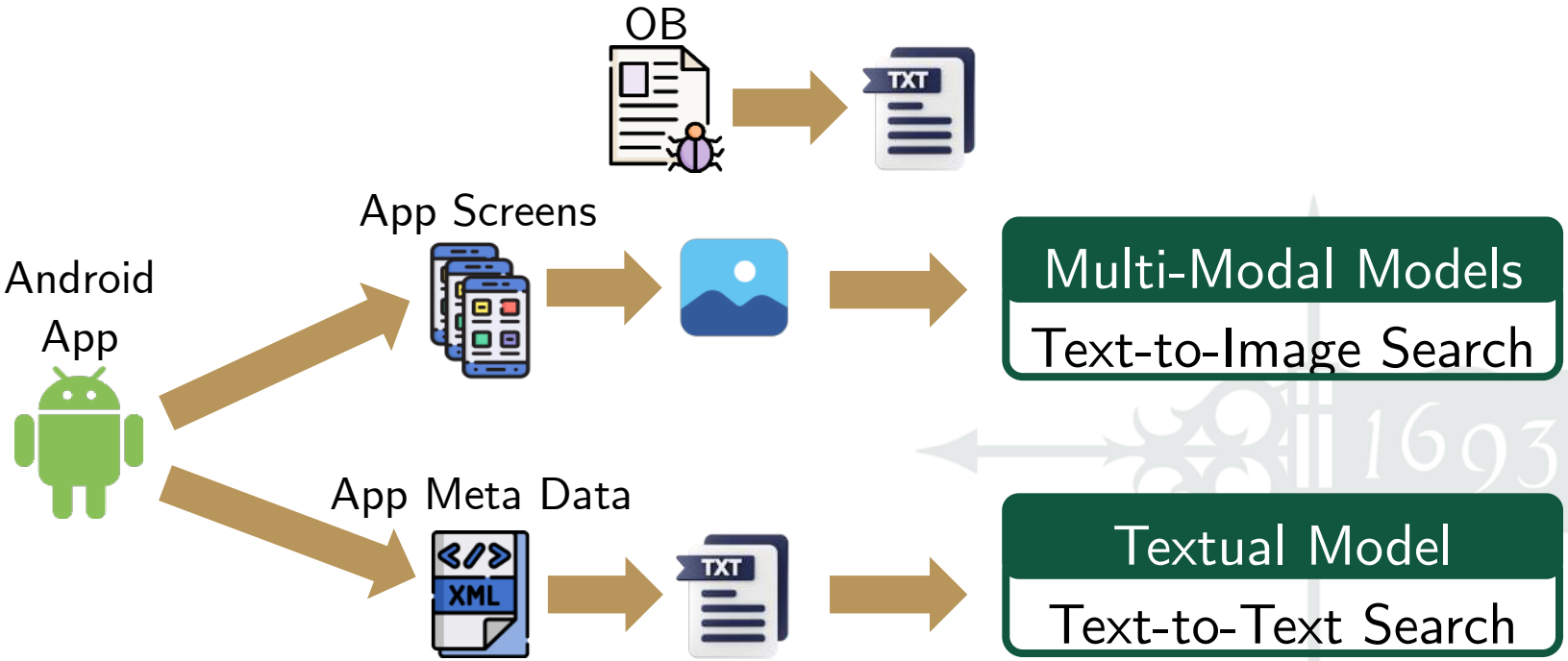
How?

Information Sources for Buggy UI Localization



How?

Information Sources for Buggy UI Localization



Study Goal & RQs



Study Goal & RQs

Goal

Investigate the Feasibility of Buggy UI Localization Using Existing Retrieval Approaches



Study Goal & RQs

Goal

Investigate the Feasibility of Buggy UI Localization Using Existing Retrieval Approaches

RQs

1. Effectiveness of Approaches for Screen Localization (SL)?
2. Effectiveness of Approaches for Component Localization (CL)?
3. Effectiveness of Approaches for Different Query Quality?

Retrieval Approaches



Retrieval Approaches



Unsupervised TF-IDF-Based Textual Model

Retrieval Approaches



Unsupervised TF-IDF-Based Textual Model

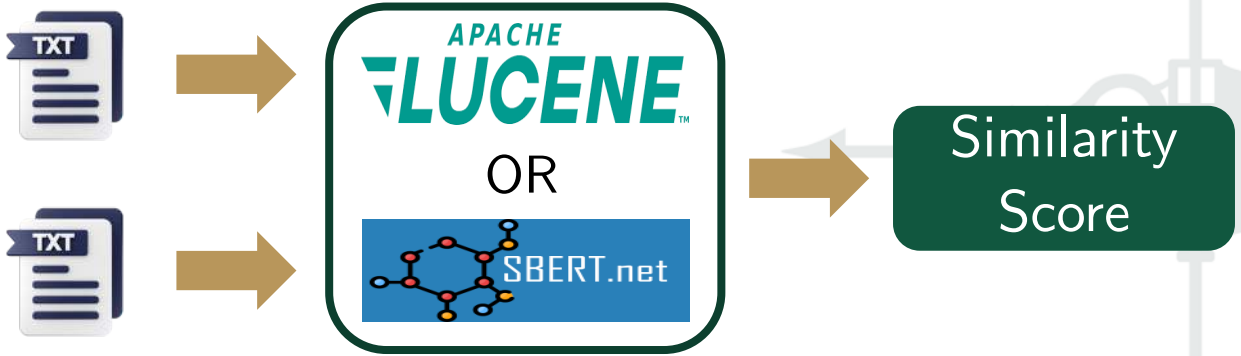


Neural BERT-Based Textual Language Model

Retrieval Approaches

Unsupervised TF-IDF-Based Textual Model

Neural BERT-Based Textual Language Model



Retrieval Approaches



Retrieval Approaches



CLIP Neural Multi-modal Vision-Language Model

Retrieval Approaches


CLIP Neural Multi-modal Vision-Language Model

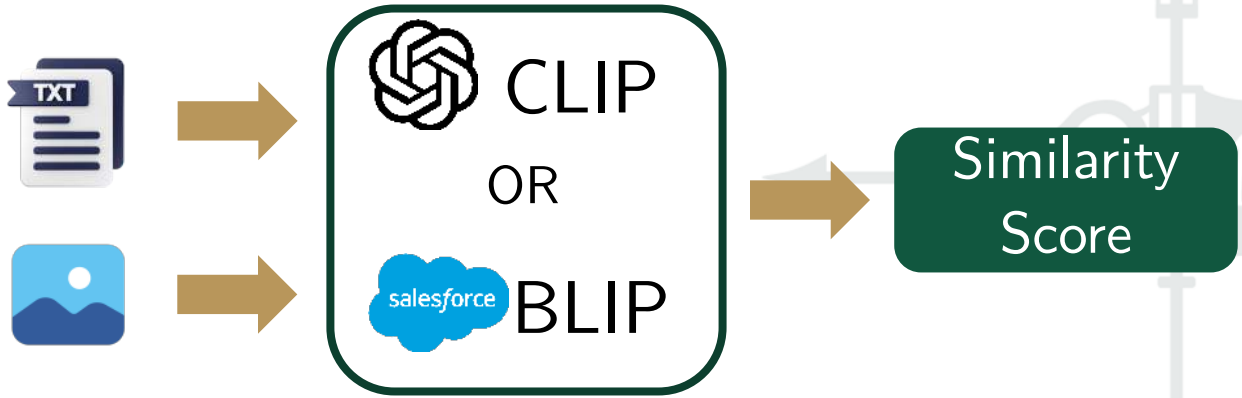


BLIP Neural Multi-modal Vision-Language Model

Retrieval Approaches

CLIP — Neural Multi-modal Vision-Language Model

BLIP — Neural Multi-modal Vision-Language Model



Dataset Construction



Dataset Construction

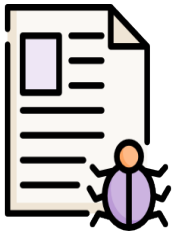


Query

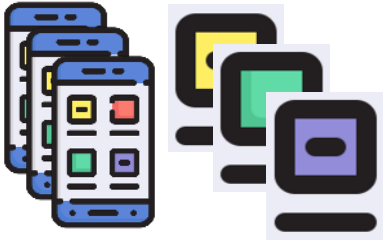
228 (SL)
254 (CL)

Dataset Construction

OB



App Screens/Comps.

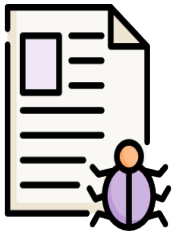


Query
228 (SL)
254 (CL)

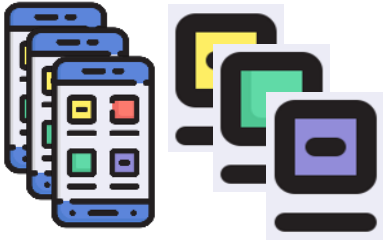
Corpus
26 Per App (SL)
17 Per Screen (CL)

Dataset Construction

OB



App Screens/Comps.



Buggy Screens/Comps.



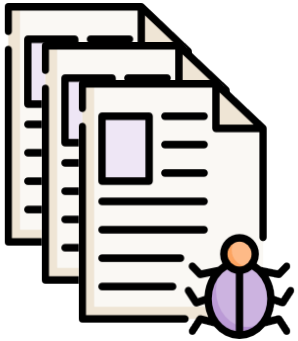
Query	Corpus	Ground Truth
228 (SL) 254 (CL)	26 Per App (SL) 17 Per Screen (CL)	2.1 Per Query (SL) 1.9 Per Query (CL)

Bug Report Selection



Bug Report Selection

87 Bug Reports[1, 2]

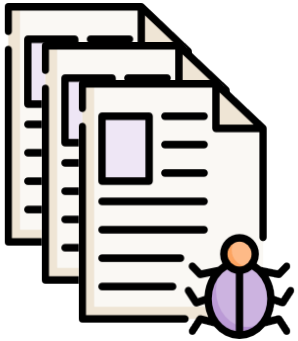


1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

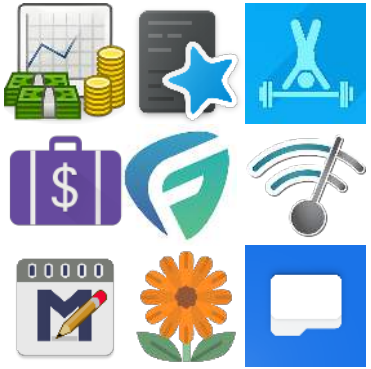
2 Wendland, Tyler, et al. "Andror2: A dataset of manually-reproduced bug reports for android apps." MSR'21.

Bug Report Selection

87 Bug Reports[1, 2]



39 Android Apps

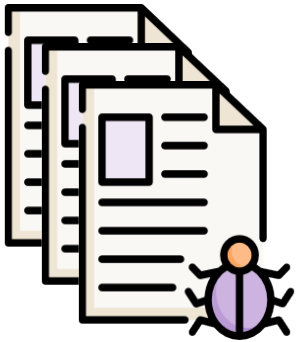


1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

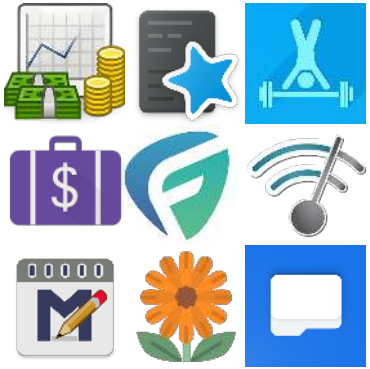
2 Wendland, Tyler, et al. "Andror2: A dataset of manually-reproduced bug reports for android apps." MSR'21.

Bug Report Selection

87 Bug Reports[1, 2]



39 Android Apps



Various Bug Types

- UI Cosmetic Issue
- App Crash
- Output Problem

1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

2 Wendland, Tyler, et al. "Andror2: A dataset of manually-reproduced bug reports for android apps." MSR'21.

Retrieval Corpus Collection



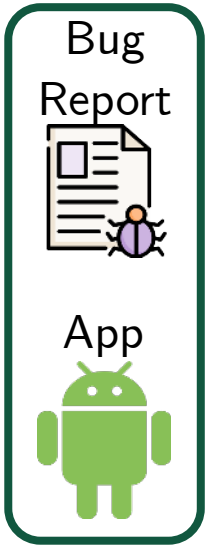
Retrieval Corpus Collection

Collect screenshots & metadata of UI screens/components



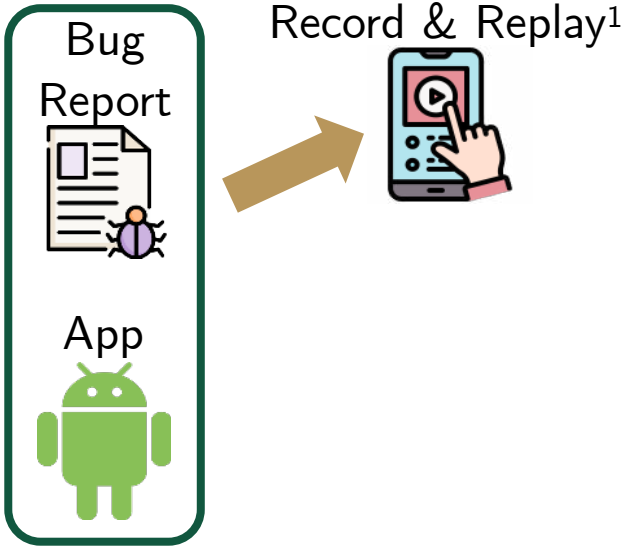
Retrieval Corpus Collection

Collect screenshots & metadata of UI screens/components



Retrieval Corpus Collection

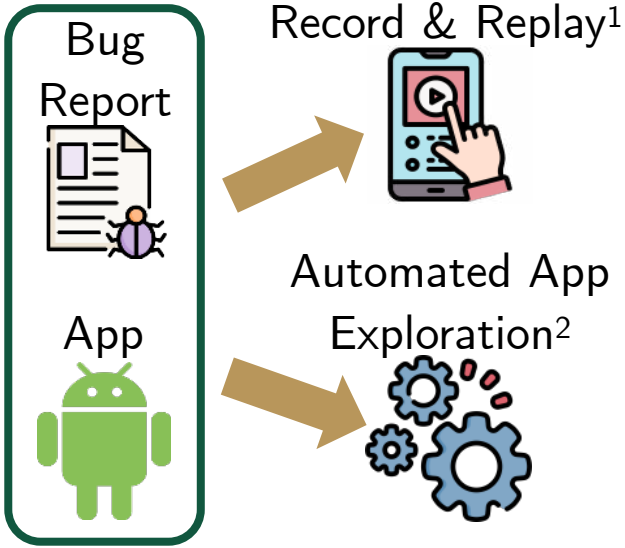
Collect screenshots & metadata of UI screens/components



1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

Retrieval Corpus Collection

Collect screenshots & metadata of UI screens/components

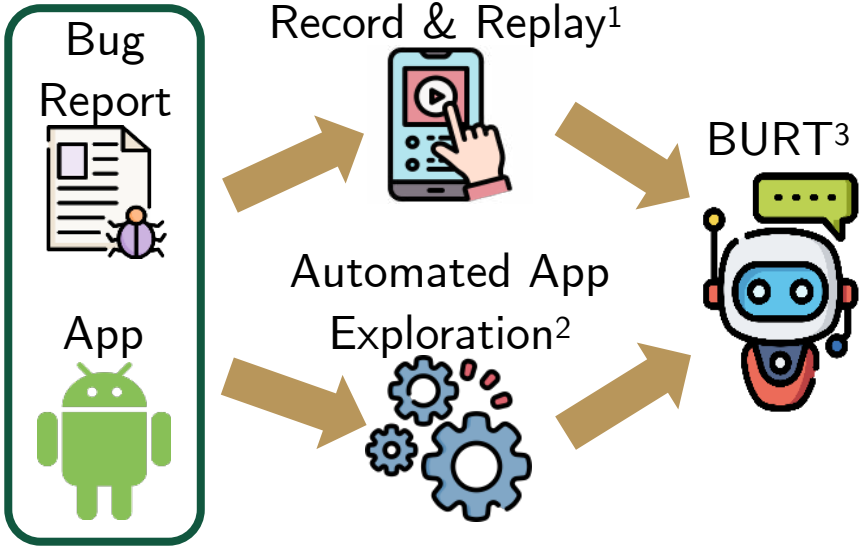


1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

2 Moran, Kevin, et al. "Crashscope: A practical tool for automated testing of android applications." ICSE'17.

Retrieval Corpus Collection

Collect screenshots & metadata of UI screens/components



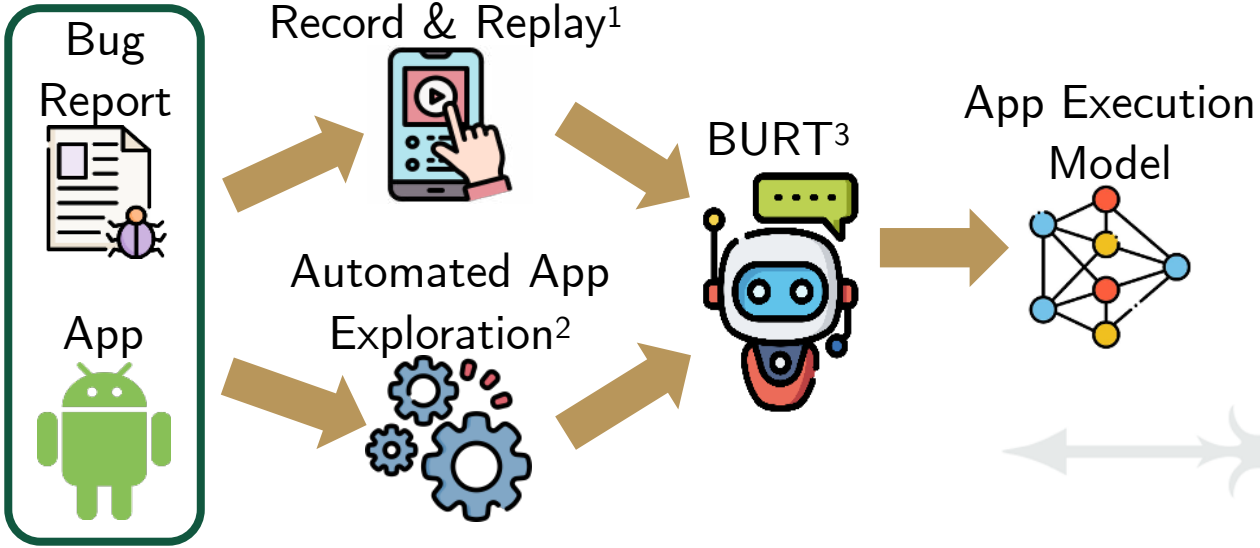
1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

2 Moran, Kevin, et al. "Crashscope: A practical tool for automated testing of android applications." ICSE'17.

3 Chaparro, Oscar, et al. "Assessing the quality of the steps to reproduce in bug reports." FSE'19.

Retrieval Corpus Collection

Collect screenshots & metadata of UI screens/components



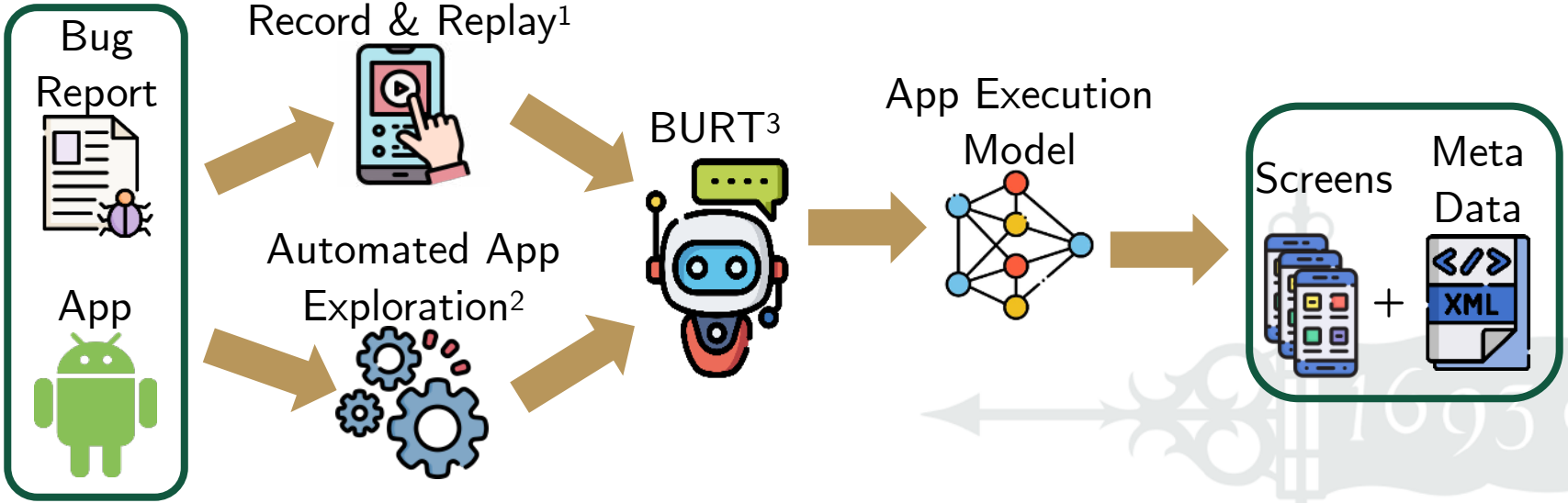
1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

2 Moran, Kevin, et al. "Crashscope: A practical tool for automated testing of android applications." ICSE'17.

3 Chaparro, Oscar, et al. "Assessing the quality of the steps to reproduce in bug reports." FSE'19.

Retrieval Corpus Collection

Collect screenshots & metadata of UI screens/components



1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

2 Moran, Kevin, et al. "Crashscope: A practical tool for automated testing of android applications." ICSE'17.

3 Chaparro, Oscar, et al. "Assessing the quality of the steps to reproduce in bug reports." FSE'19.

Ground Truth Construction



Ground Truth Construction

For each OB, identify the buggy UI screens & components



Ground Truth Construction

For each OB, identify the buggy UI screens & components

4 Annotators



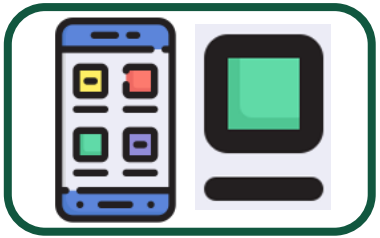
Ground Truth Construction

For each OB, identify the buggy UI screens & components

4 Annotators



Buggy UI Screen & Component



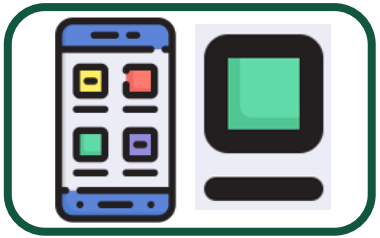
Ground Truth Construction

For each OB, identify the buggy UI screens & components

4 Annotators



Buggy UI Screen & Component



OB Quality Rating



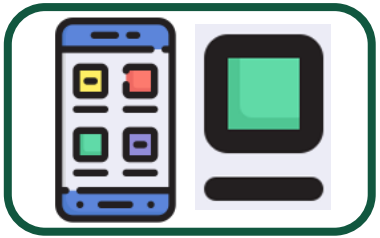
Ground Truth Construction

For each OB, identify the buggy UI screens & components

4 Annotators



Buggy UI Screen & Component



OB Quality Rating



Retrieval Difficulty Levels



EASY HARD

Screen Loc. (SL) Execution

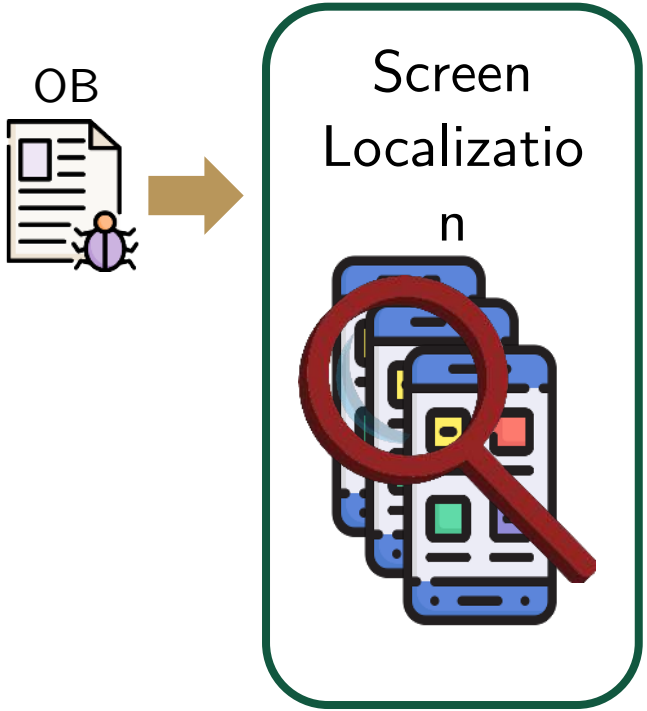


Screen Loc. (SL) Execution

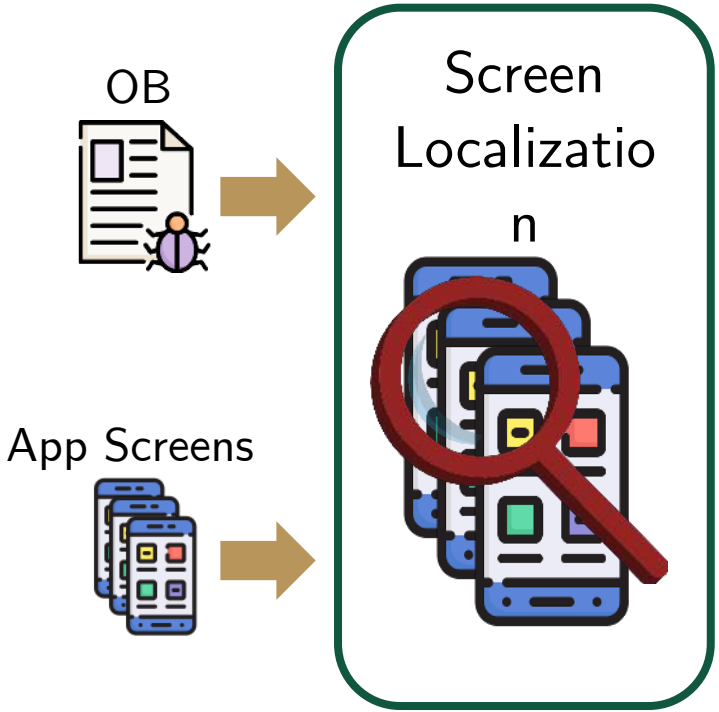
Screen Localizatio n



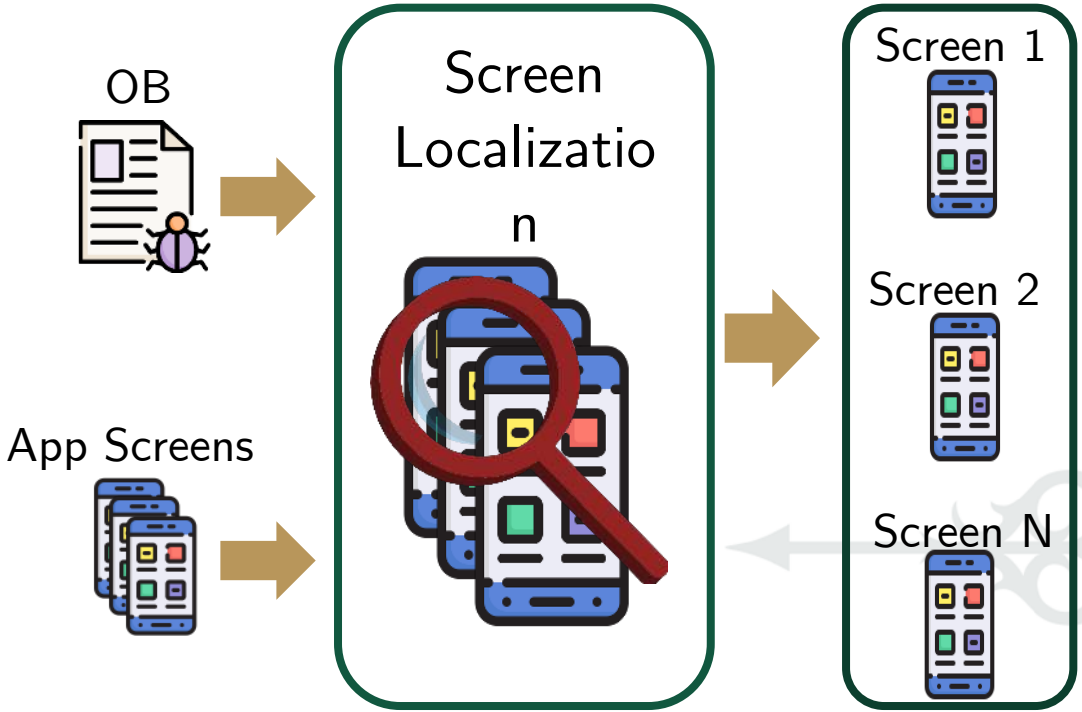
Screen Loc. (SL) Execution



Screen Loc. (SL) Execution



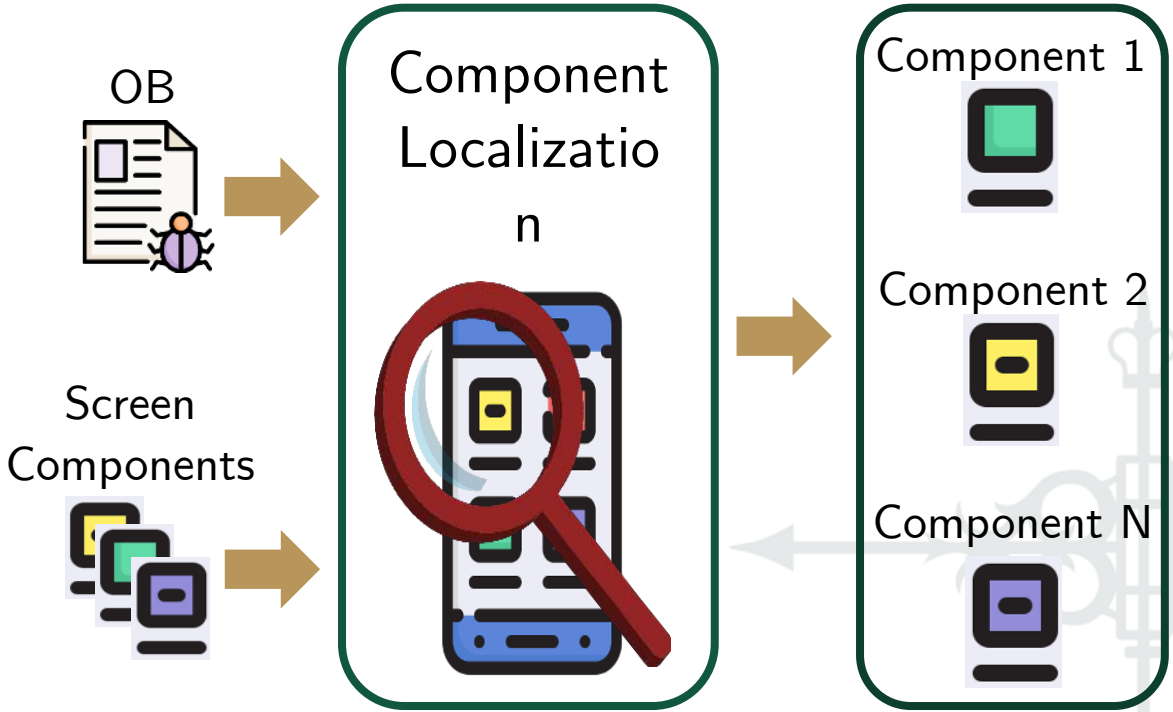
Screen Loc. (SL) Execution



Component Loc. (CL) Execution



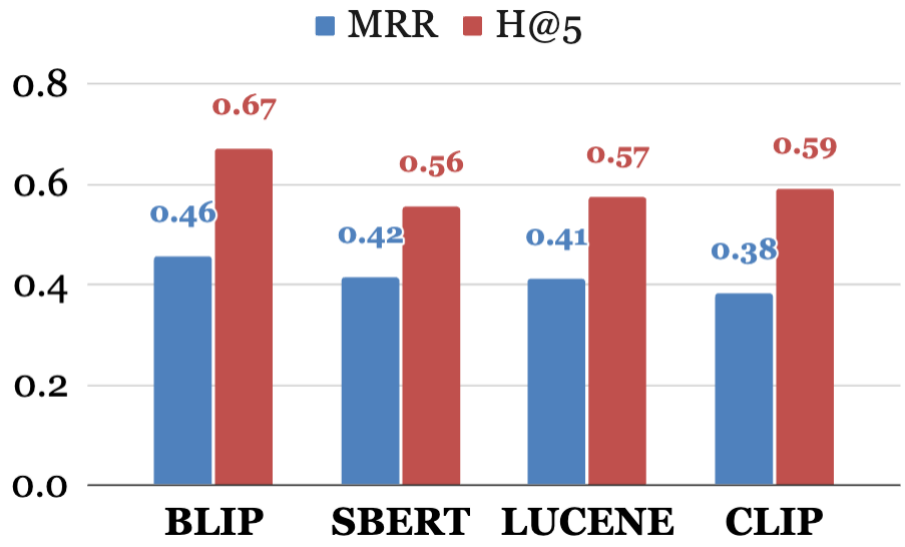
Component Loc. (CL) Execution



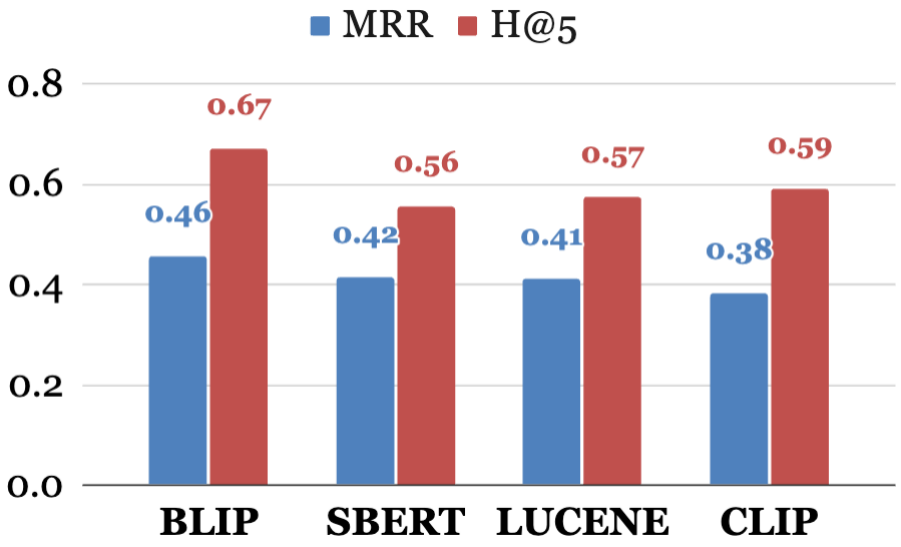
RQ1: Screen Localization (SL) Results



RQ1: Screen Localization (SL) Results



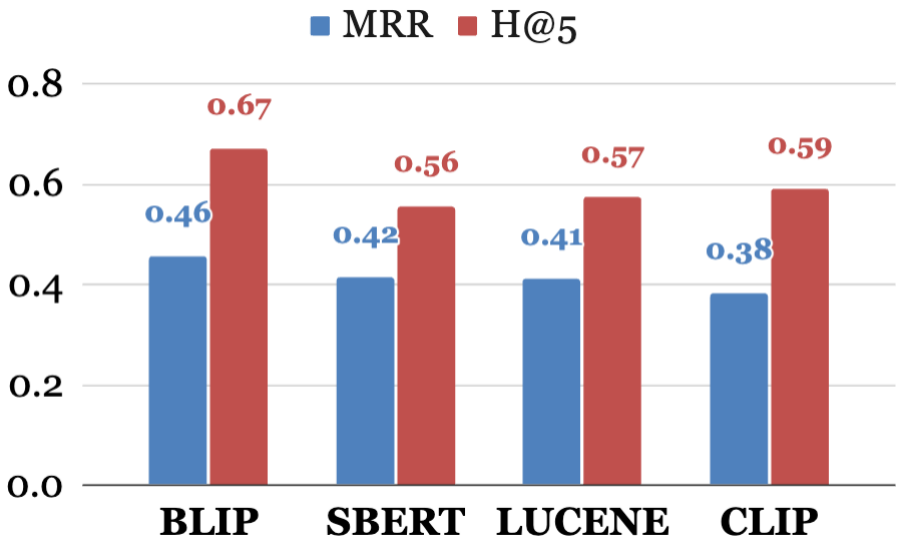
RQ1: Screen Localization (SL) Results



RQ1: Screen Localization (SL) Results

Best SL Model

BLIP



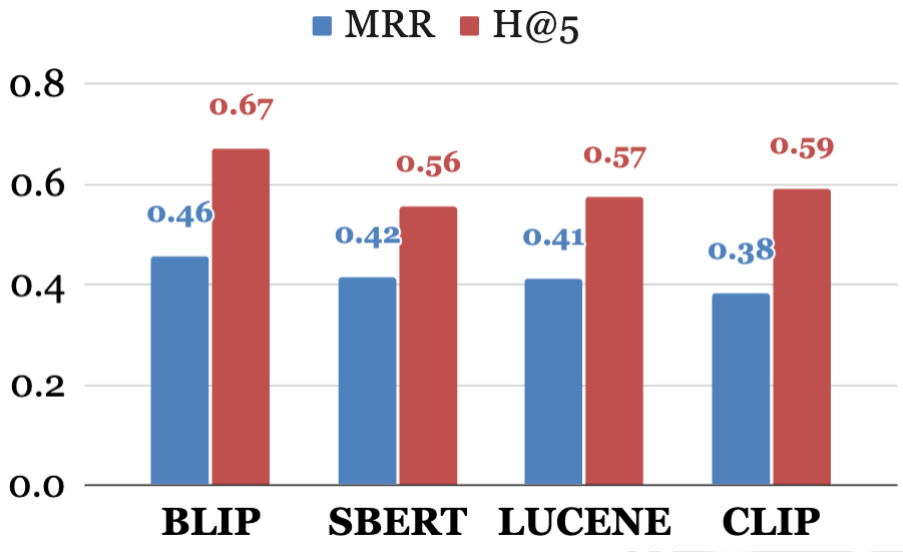
RQ1: Screen Localization (SL) Results

Best SL Model

BLIP

Outperforms SBERT

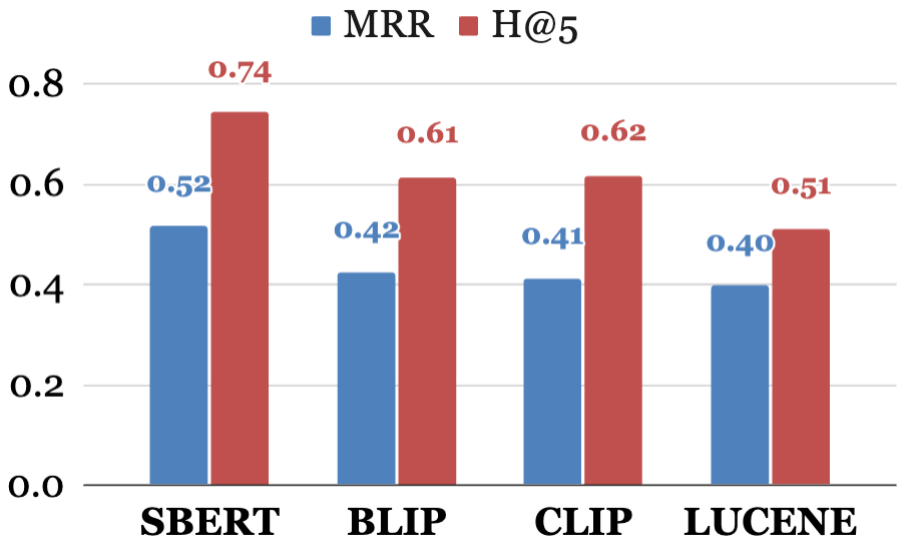
By 10.1% (MRR)
By 20.5% (H@5)



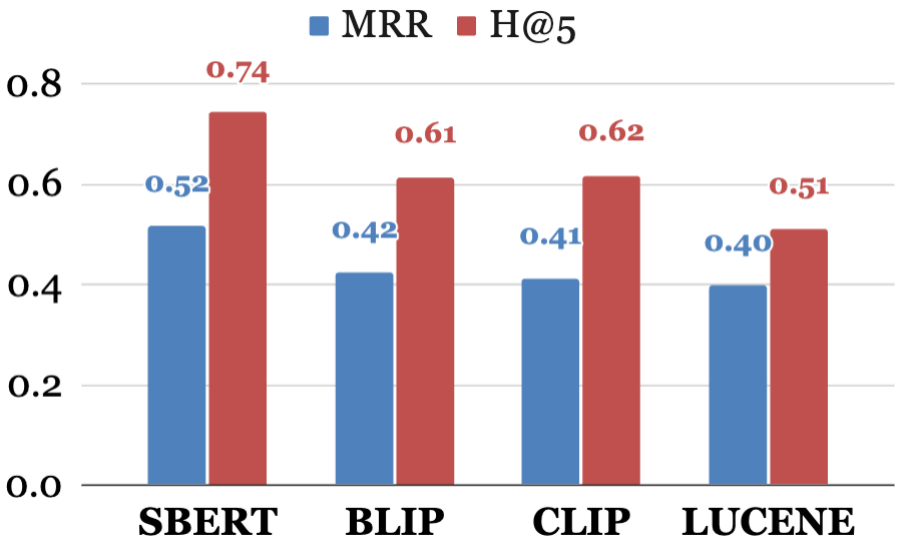
RQ2: Comp. Localization (CL) Results



RQ2: Comp. Localization (CL) Results



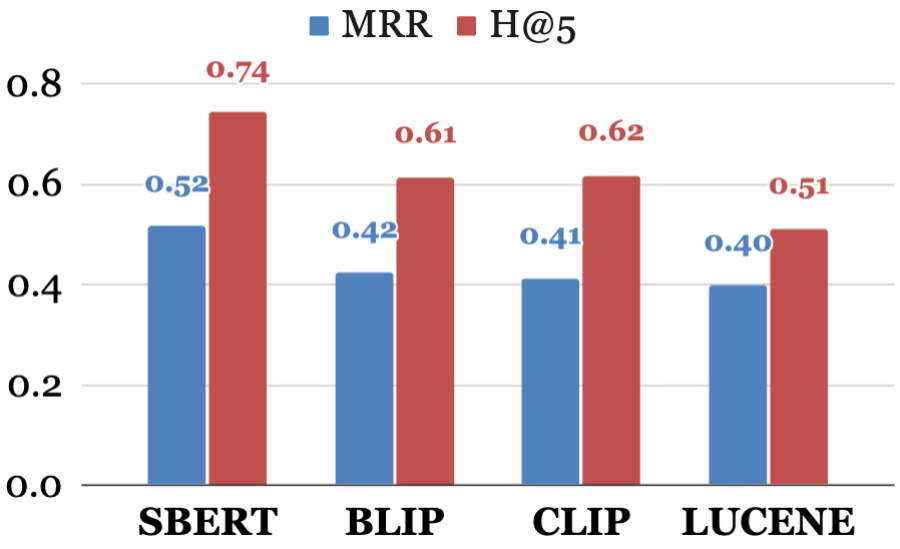
RQ2: Comp. Localization (CL) Results



RQ2: Comp. Localization (CL) Results

Best CL Model

SBERT



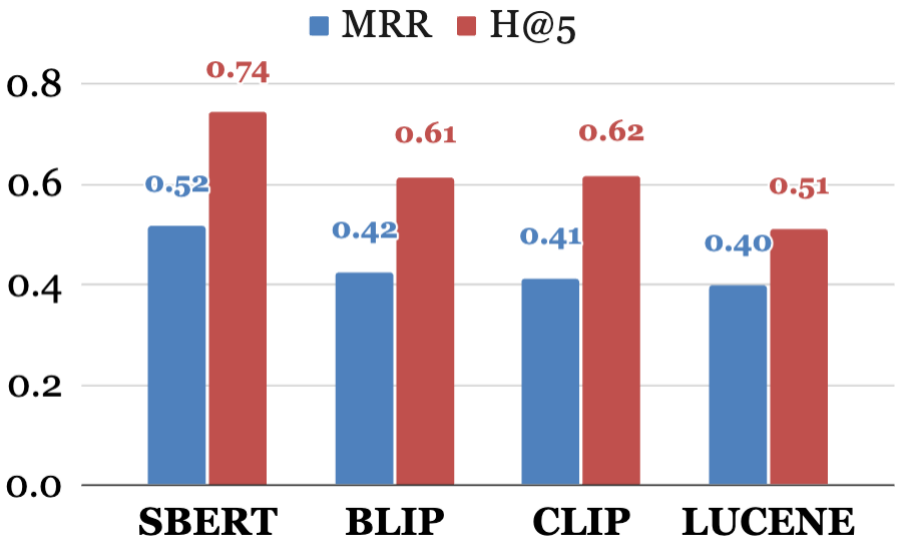
RQ2: Comp. Localization (CL) Results

Best CL Model

SBERT

Outperforms BLIP

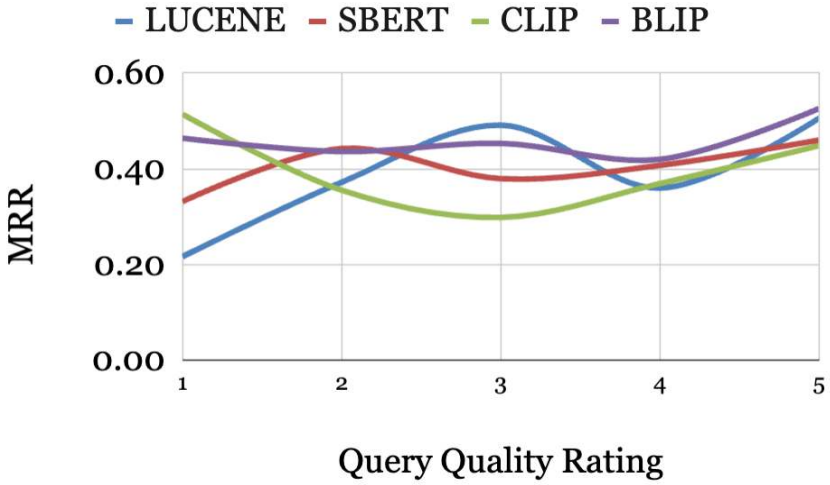
By 21.9% (MRR)
By 21.2% (H@5)



RQ3: Query Quality vs. Results

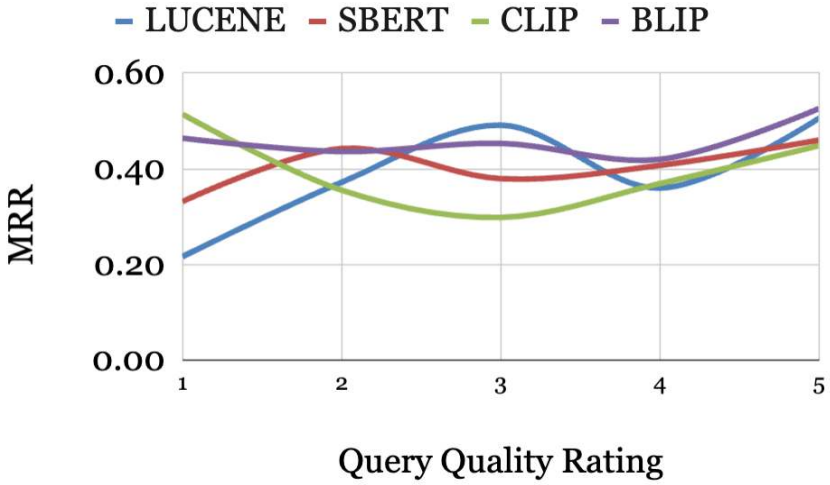


RQ3: Query Quality vs. Results

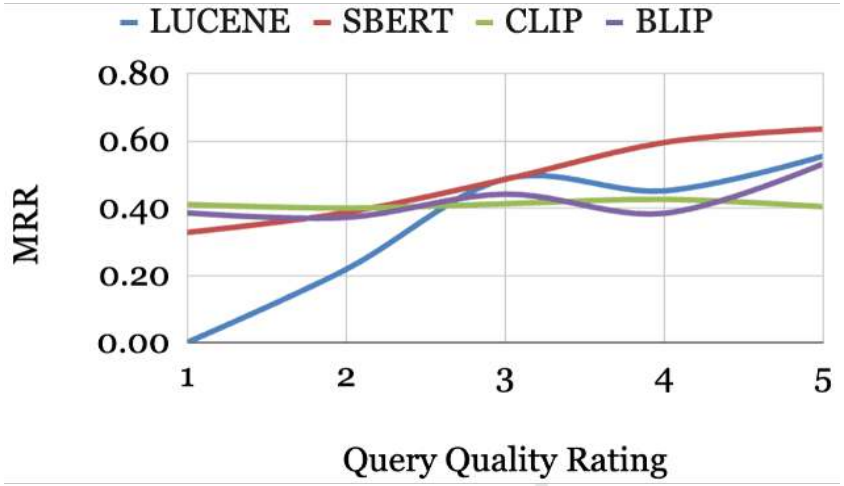


SL

RQ3: Query Quality vs. Results

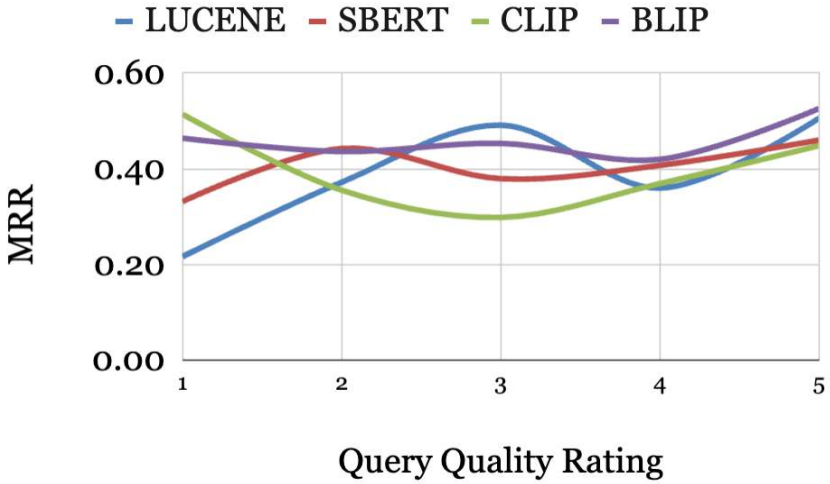


SL

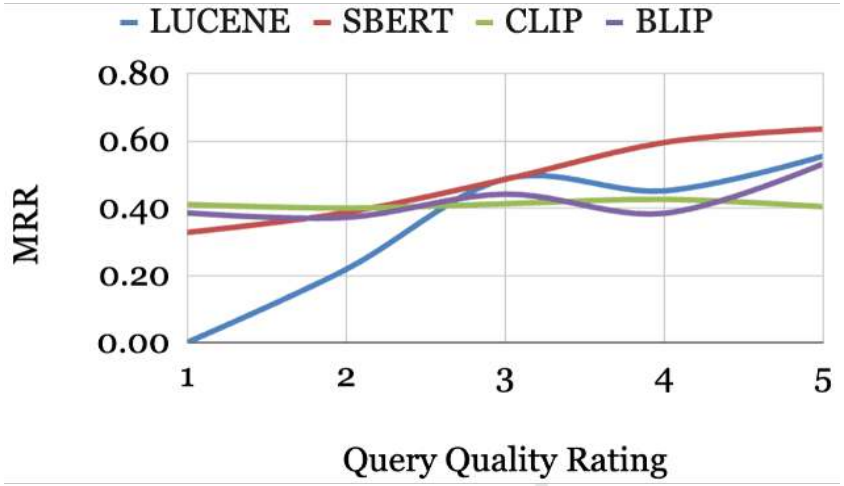


CL

RQ3: Query Quality vs. Results



SL



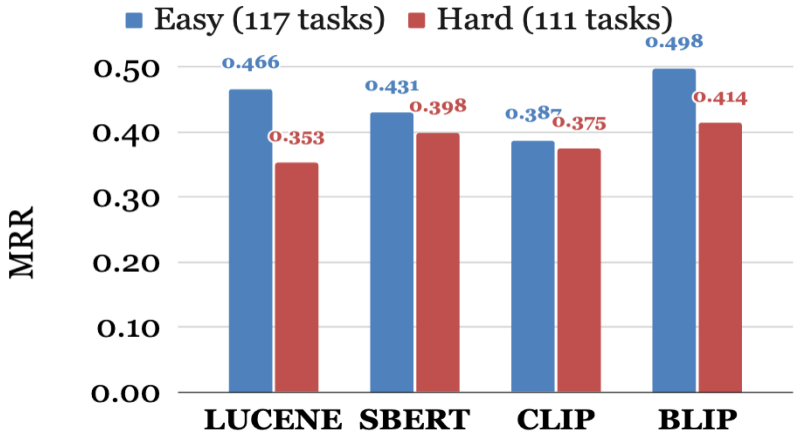
CL

All models perform best for the most informative queries (rating =

RQ3: Retrieval Difficulty vs. SL Results

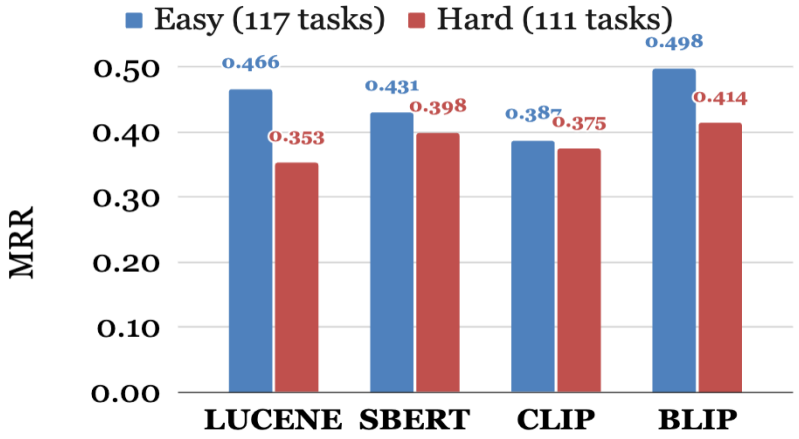


RQ3: Retrieval Difficulty vs. SL Results

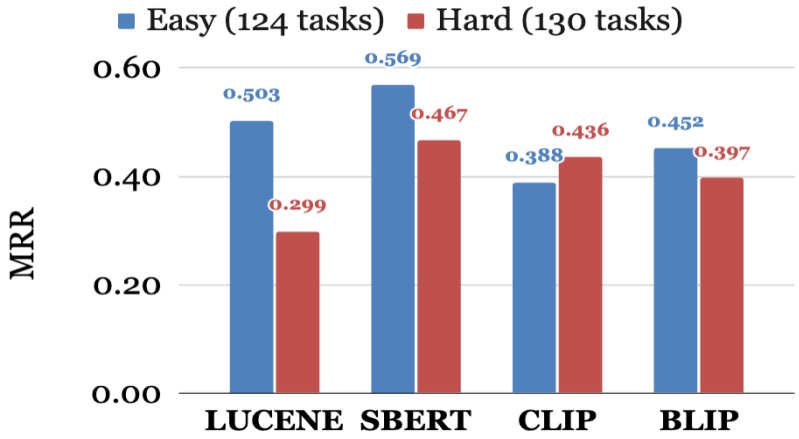


SL

RQ3: Retrieval Difficulty vs. SL Results

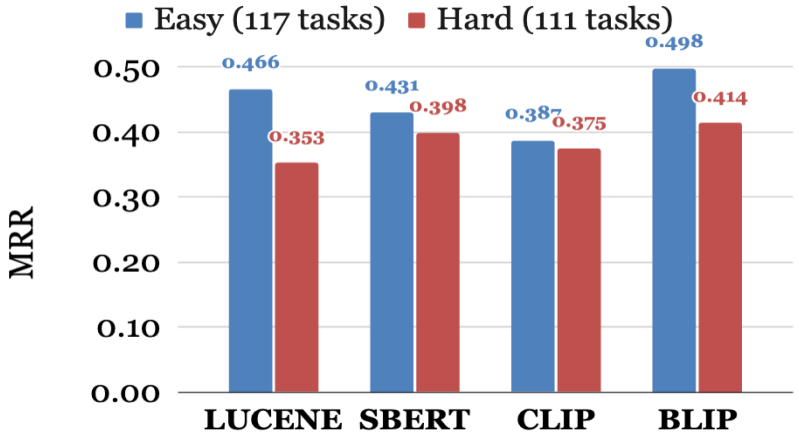


SL

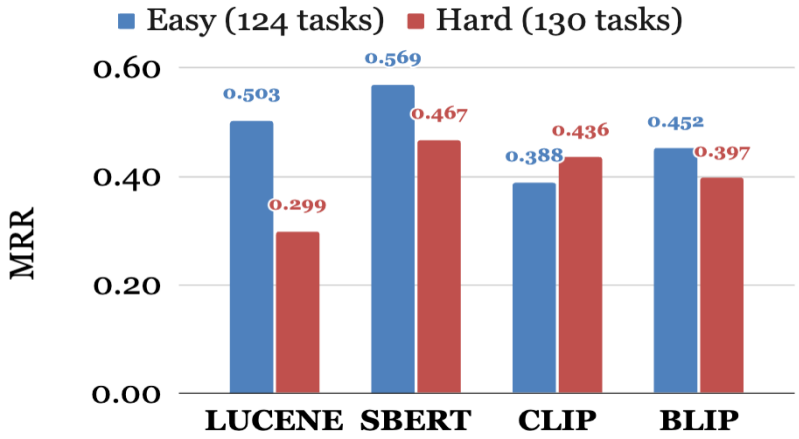


CL

RQ3: Retrieval Difficulty vs. SL Results



SL



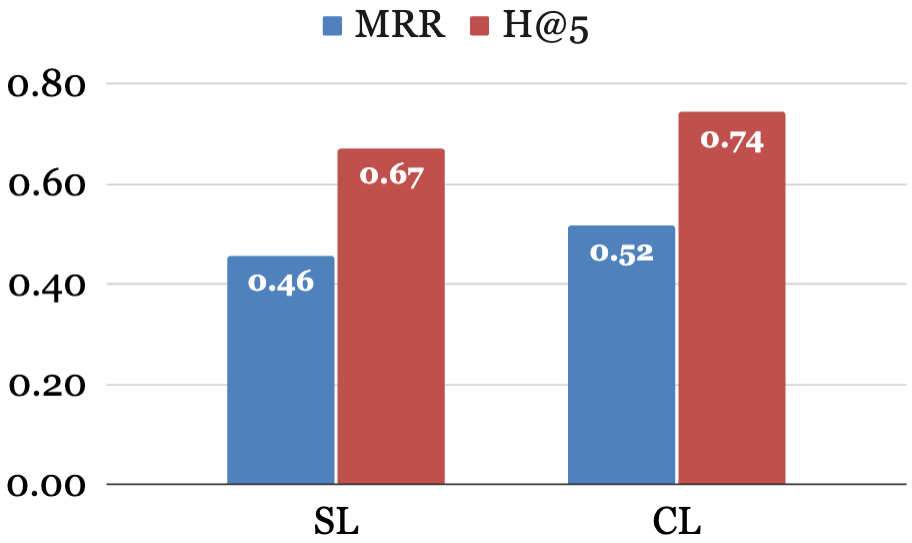
CL

All models perform higher on easy-to-retrieve tasks

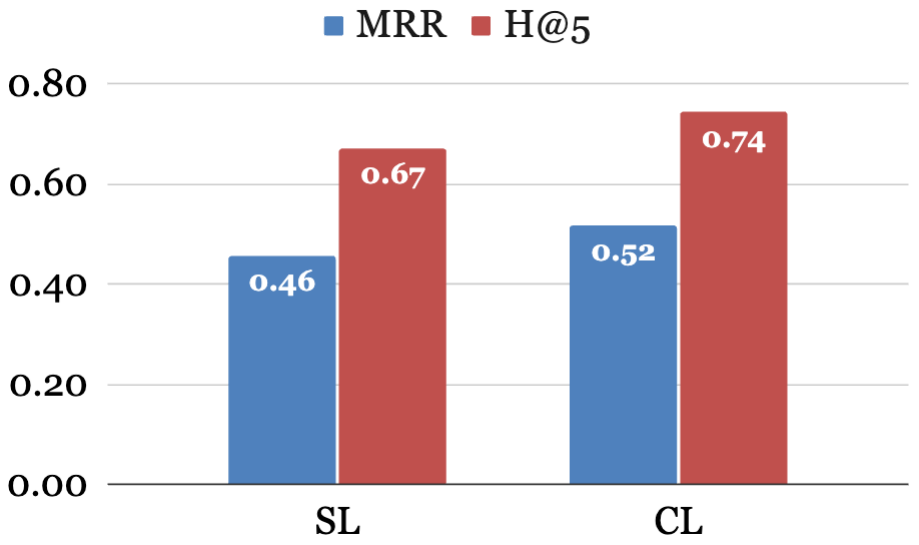
Screen vs. Component Localization



Screen vs. Component Localization

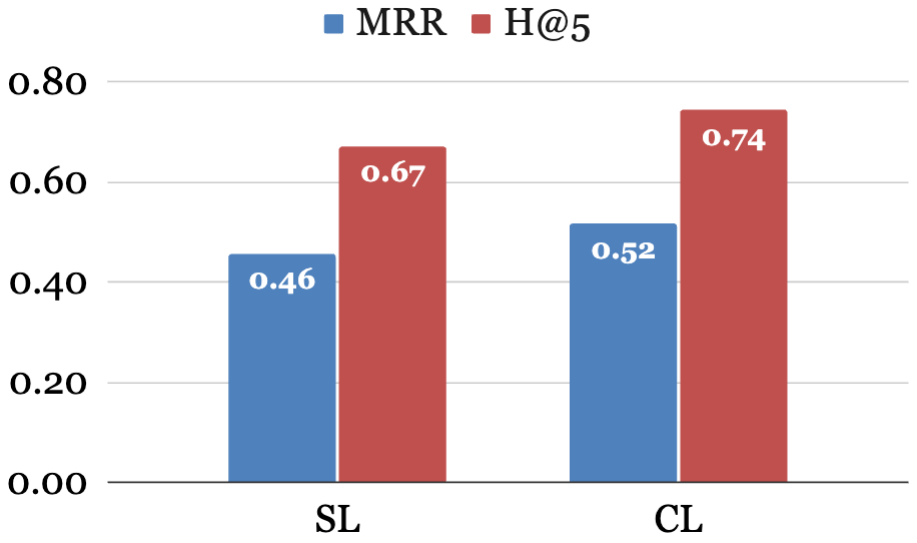


Screen vs. Component Localization



Screen vs. Component Localization

SL is more challenging

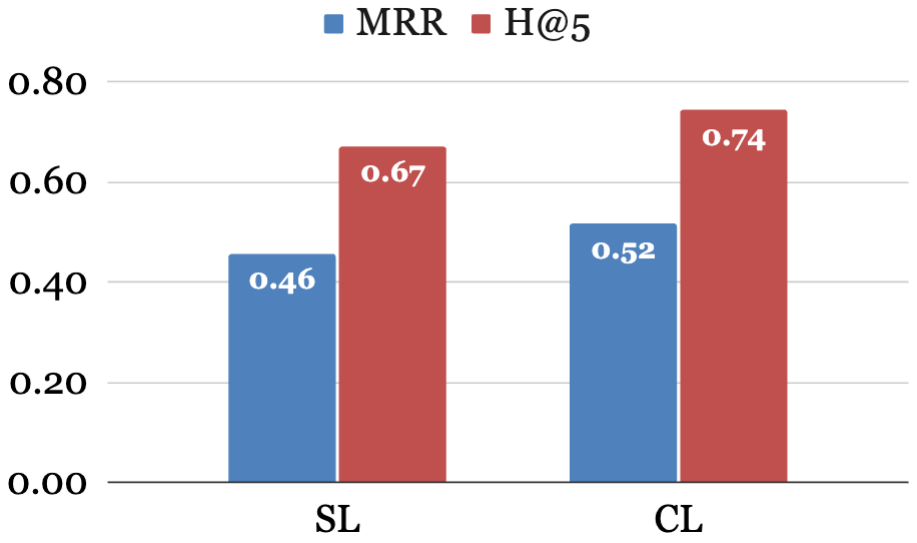


Screen vs. Component Localization

SL is more challenging

Corpus Size

26 per App vs.
17 per Screen



Screen vs. Component Localization

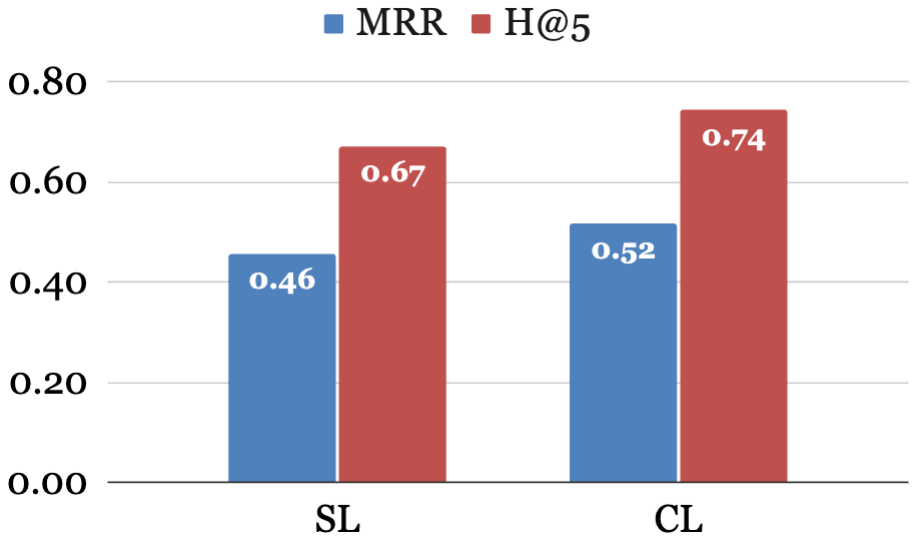
SL is more challenging

Corpus Size

26 per App vs.
17 per Screen

Scope of SL is Broader

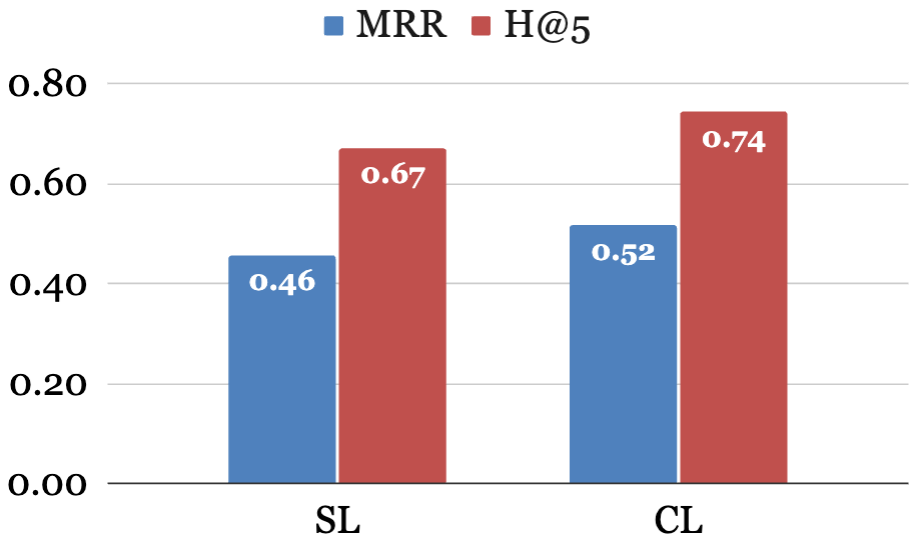
All Screens of App vs.
All comp. of a Screen



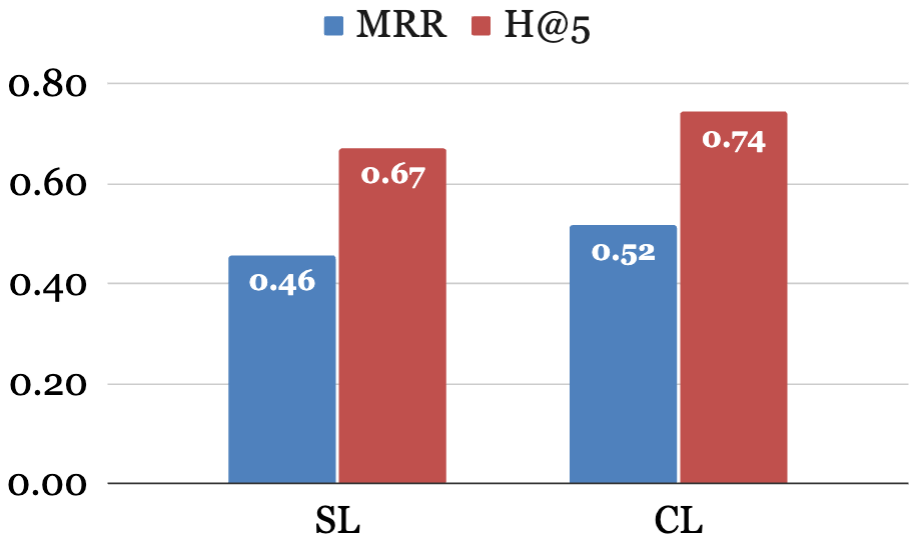
Textual vs. Multi-modal Models



Textual vs. Multi-modal Models

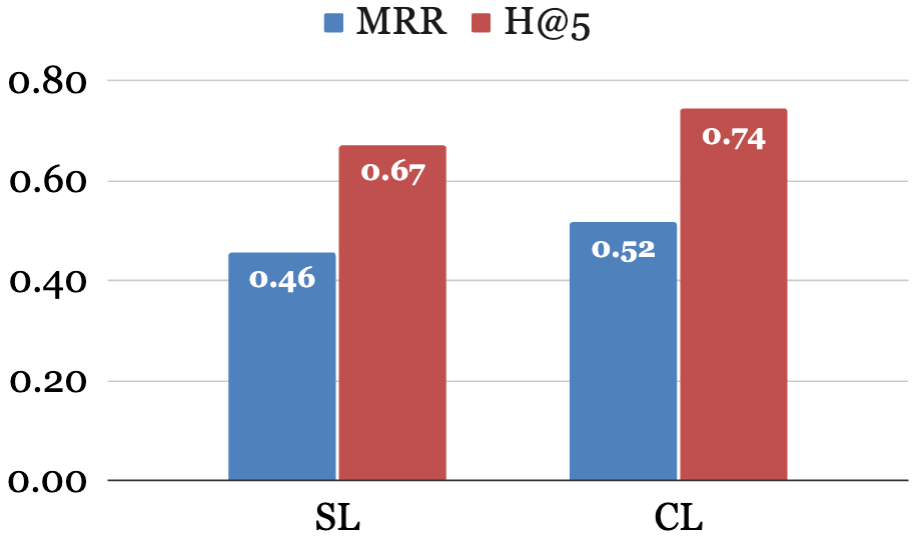


Textual vs. Multi-modal Models



BLIP

Textual vs. Multi-modal Models



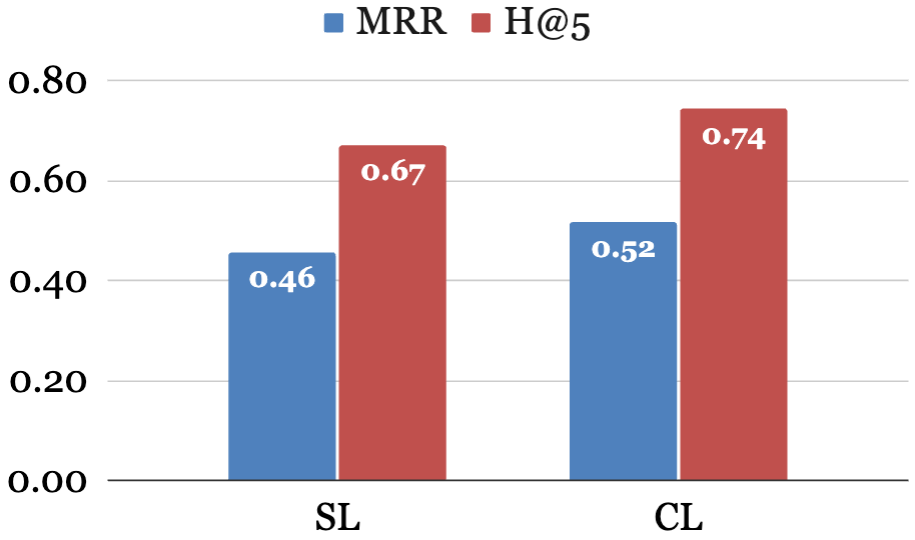
BLIP

SBERT

Textual vs. Multi-modal Models

SL: Multi-Modal Model

Model should understand full context of the screen



BLIP

SBERT

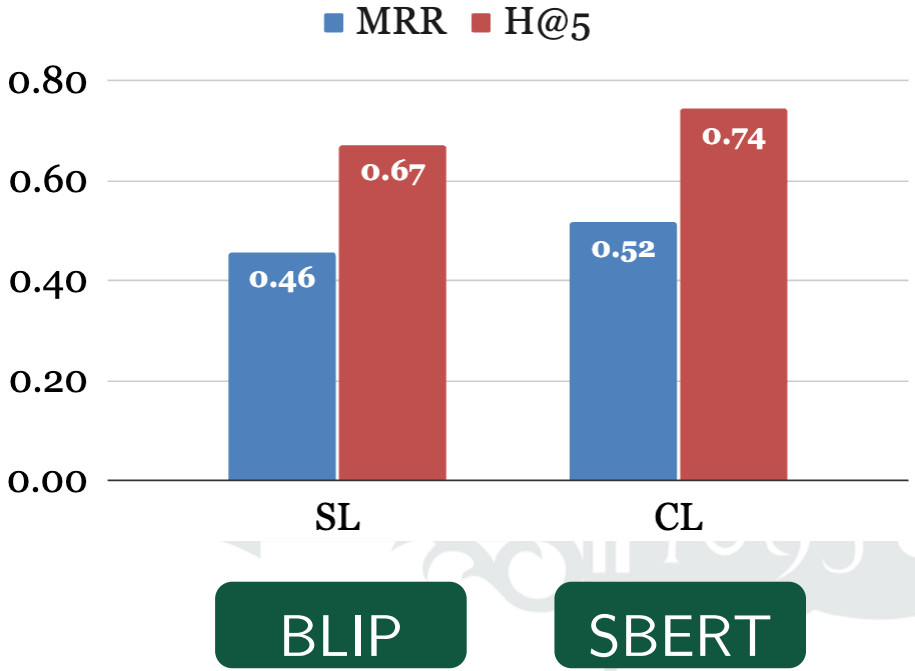
Textual vs. Multi-modal Models

SL: Multi-Modal Model

Model should understand full context of the screen

CL: Textual Model

OBs are written using component text

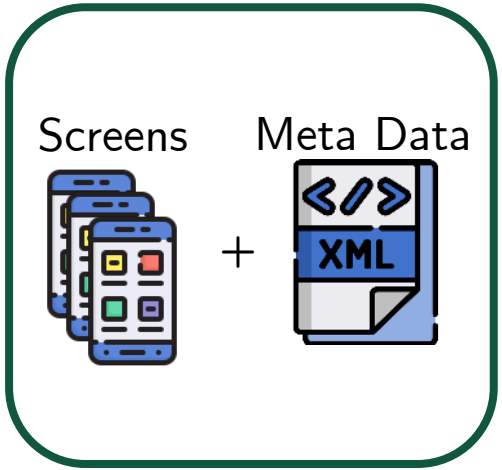


Enhanced Buggy UI Localization



Enhanced Buggy UI Localization

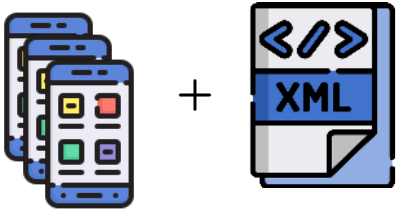
Both Visual & Textual Information



Enhanced Buggy UI Localization

Both Visual & Textual Information

Screens + Meta Data



Distinct Models for Screen & Component Localization

Screen Localizatio



Component Localizatio



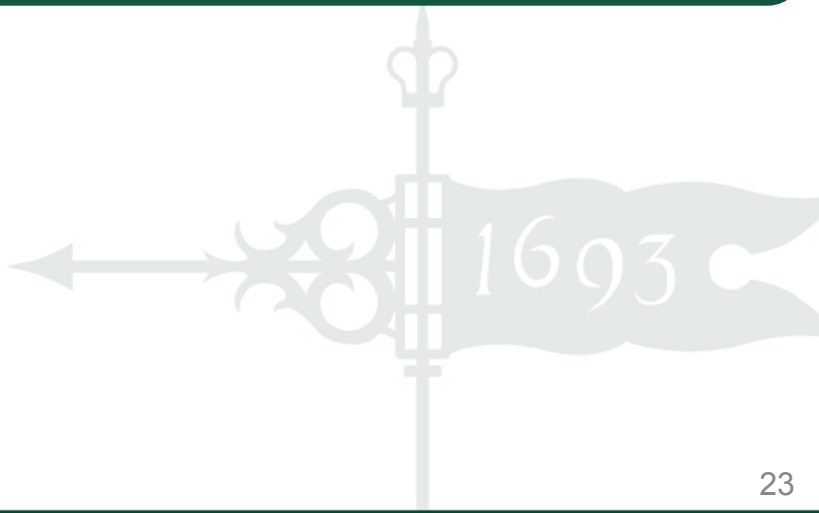
Usefulness of Buggy UI Localization



Usefulness of Buggy UI Localization

Goal

Apply Buggy UI Localization to automate IR-Based Buggy Code Localization



Usefulness of Buggy UI Localization

Goal

Apply Buggy UI Localization to automate IR-Based Buggy Code Localization

RQs

4. Can the identified buggy UI screens by BLIP lead to improved buggy code localization?

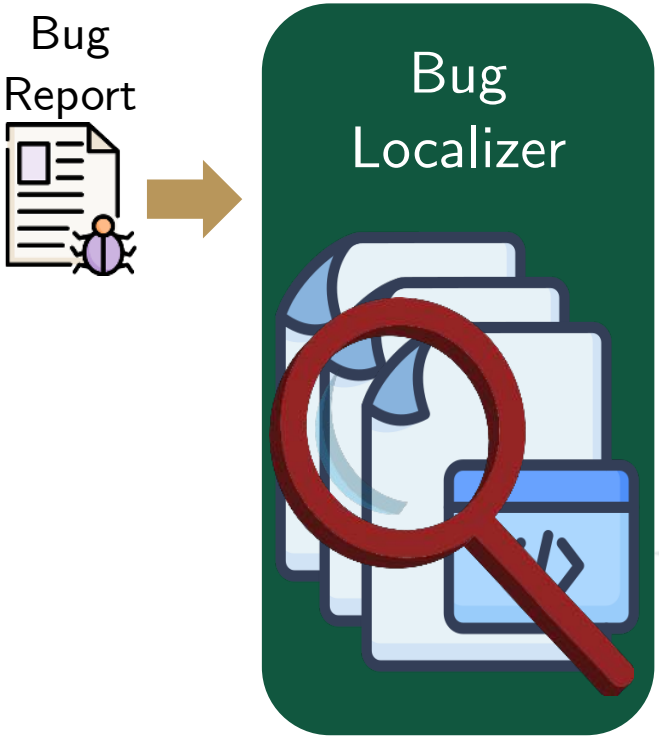
Buggy Code Localization



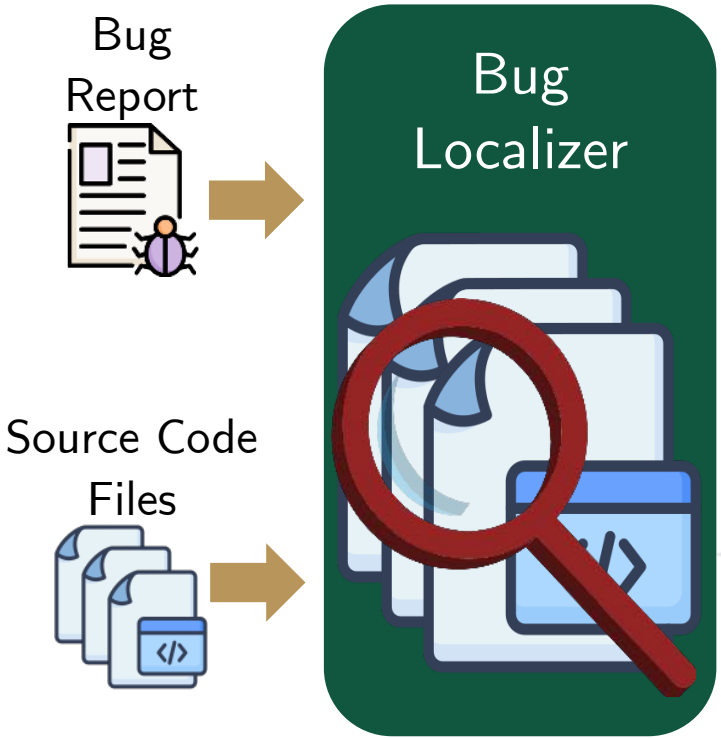
Buggy Code Localization



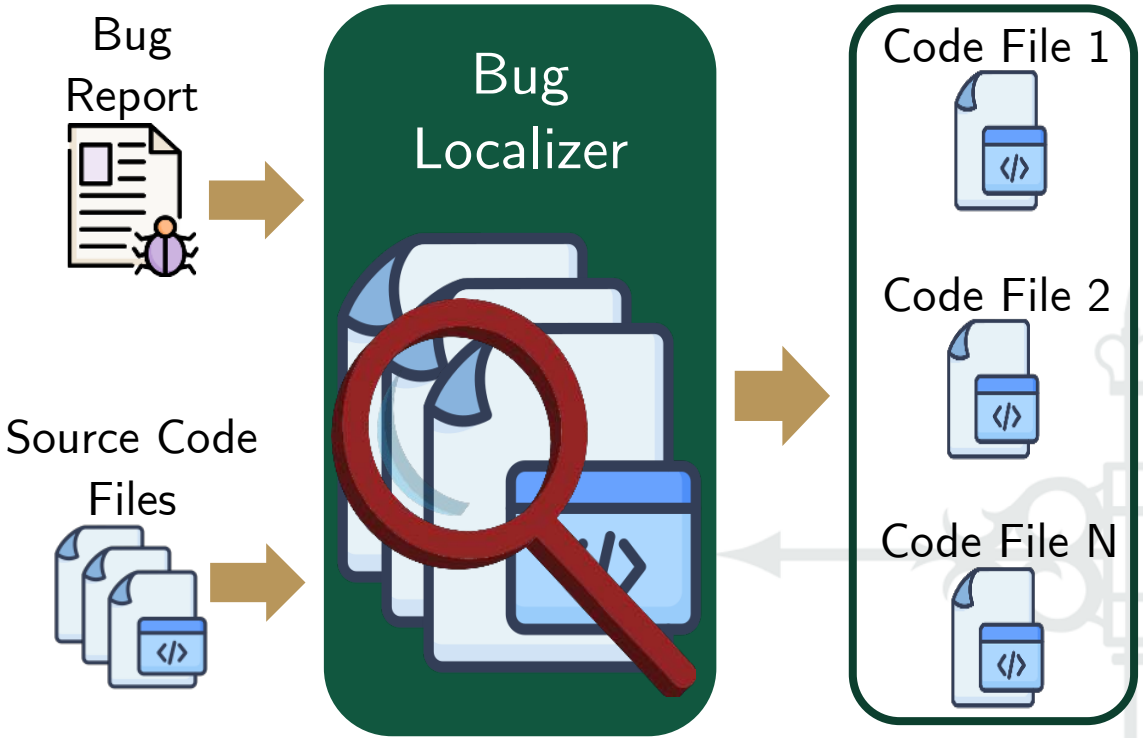
Buggy Code Localization



Buggy Code Localization



Buggy Code Localization

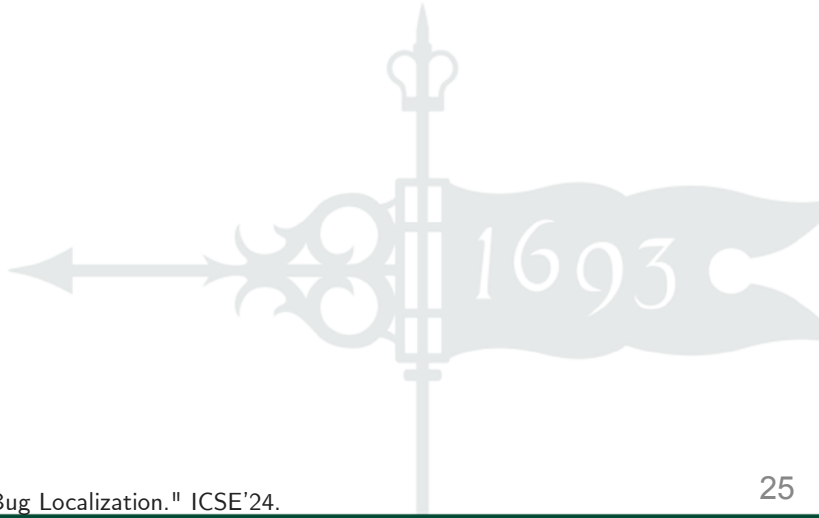


Integrating Buggy UI and Code Localization



Integrating Buggy UI and Code Localization

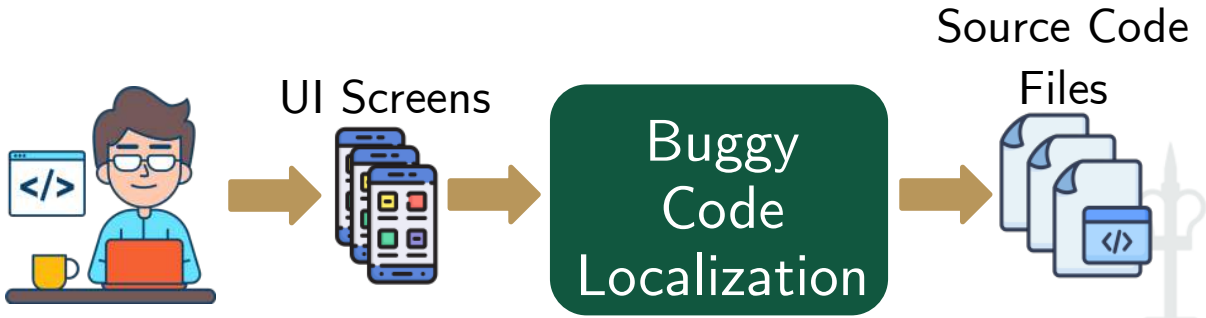
UI interaction data can improve IR-based buggy code localizers[1]



1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

Integrating Buggy UI and Code Localization

UI interaction data can improve IR-based buggy code localizers[1]



1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

Integrating Buggy UI and Code Localization

UI interaction data can improve IR-based buggy code localizers[1]



Our goal is to eliminate the manual effort of developers

1 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE'24.

Integrating Buggy UI & Code Loc.



Integrating Buggy UI & Code Loc.

End-to-End Automated Approach



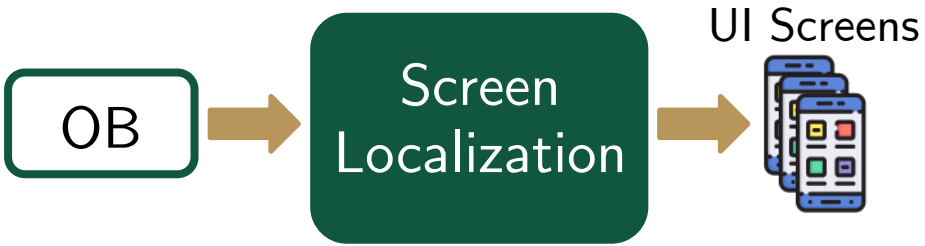
Integrating Buggy UI & Code Loc.

End-to-End Automated Approach

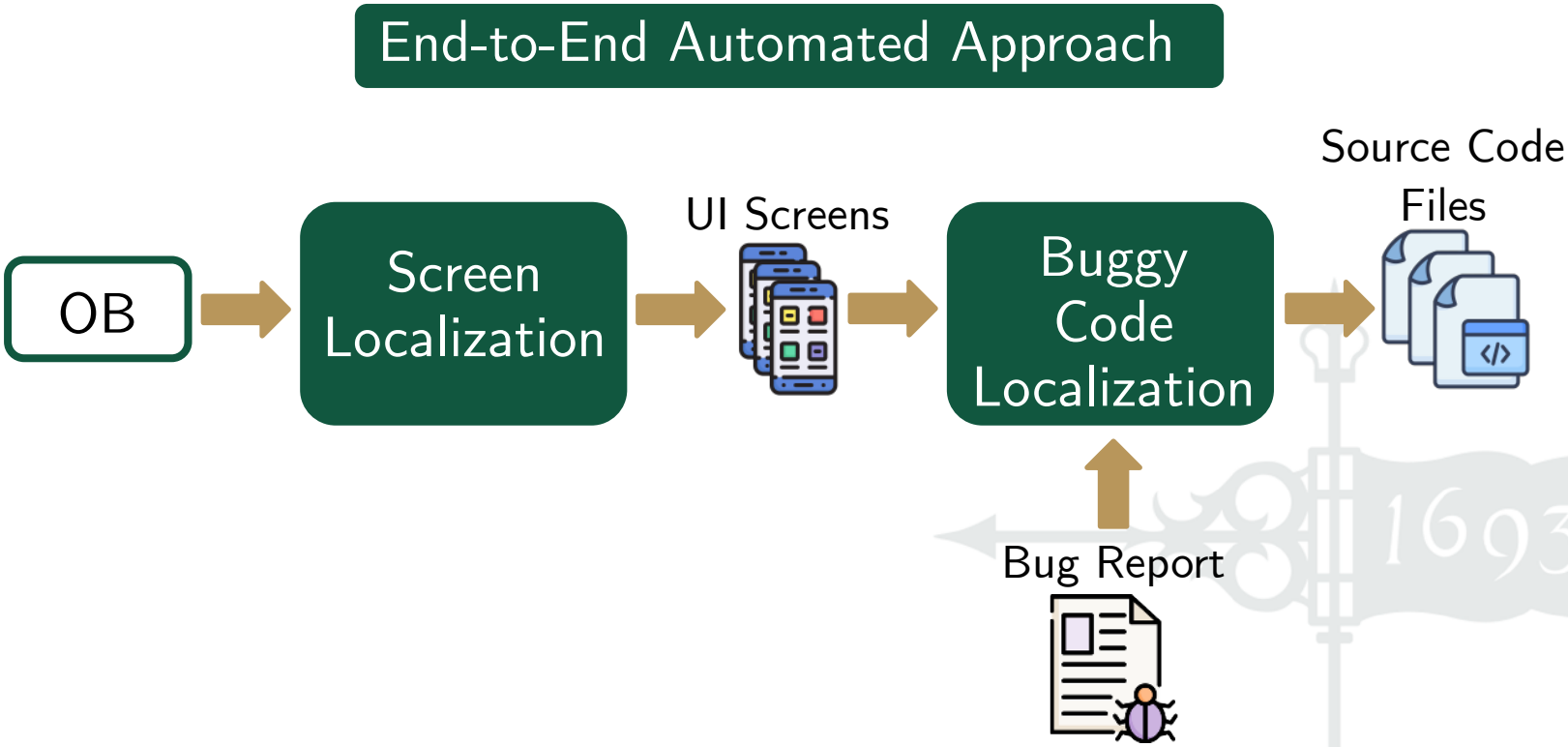


Integrating Buggy UI & Code Loc.

End-to-End Automated Approach



Integrating Buggy UI & Code Loc.



Integrating Buggy UI and Code Localization



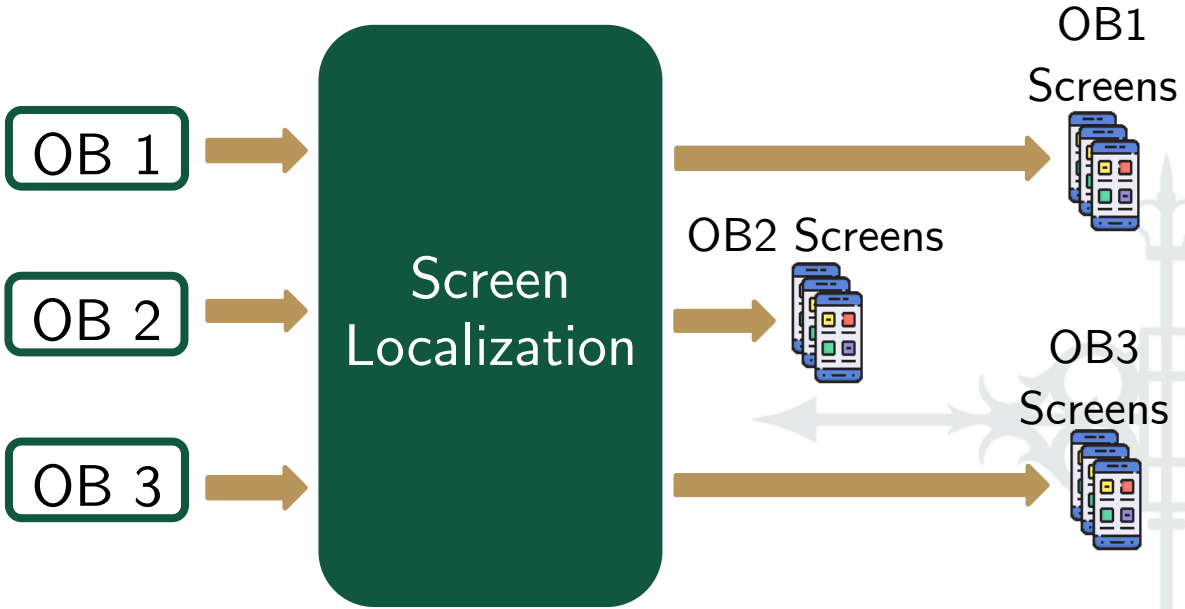
Integrating Buggy UI and Code Localization

Challenge: One bug report can have multiple OB descriptions



Integrating Buggy UI and Code Localization

Challenge: One bug report can have multiple OB descriptions



Integrating Buggy UI and Code Localization



Integrating Buggy UI and Code Localization

Which set of screens should be used for Buggy Code Localization?



Integrating Buggy UI and Code Localization

Which set of screens should be used for Buggy Code Localization?

Pipeline#1

ConcatOB

OB 1
+
OB 2
+
OB 3

Integrating Buggy UI and Code Localization

Which set of screens should be used for Buggy Code Localization?

Pipeline#1

ConcatOB

OB 1
+
OB 2
+
OB 3

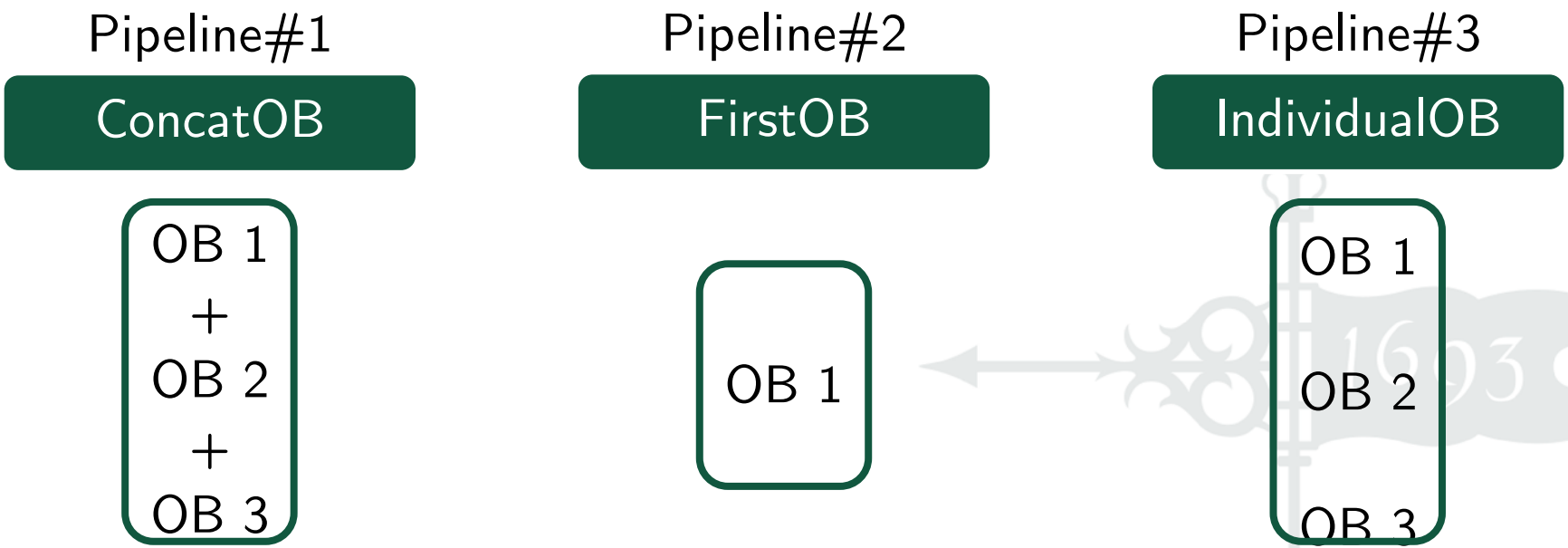
Pipeline#2

FirstOB

OB 1

Integrating Buggy UI and Code Localization

Which set of screens should be used for Buggy Code Localization?



Buggy Code Localization Execution



Buggy Code Localization Execution

Buggy UI Localization Pipelines

ConcatOB | TitleOB | IndividualO

Buggy Code Localization Execution

Buggy UI Localization Pipelines

ConcatOB | TitleOB | IndividualO

Approaches

LUCENE[1] | BugLocator[2]

1 Gospodnetic, Otis, Erik Hatcher, and Michael McCandless. Lucene in action. Simon and Schuster, 2010.

2 Zhou, Jian, et al. "Where should the bugs be fixed? more accurate information retrieval-based bug localization based on bug reports." ICSE'12.

Buggy Code Localization Execution

Buggy UI Localization Pipelines: ConcatOB | TitleOB | IndividualO

Approaches: LUCENE[1] | BugLocator[2]

Evaluation Dataset: 79 Bug Reports from Mahmud et al.[3]

1 Gospodnetic, Otis, Erik Hatcher, and Michael McCandless. Lucene in action. Simon and Schuster, 2010.

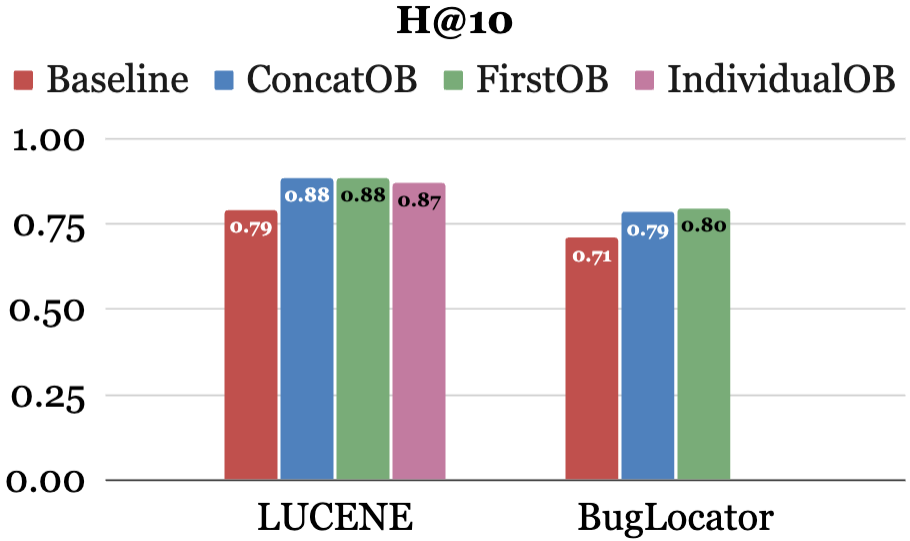
2 Zhou, Jian, et al. "Where should the bugs be fixed? more accurate information retrieval-based bug localization based on bug reports." ICSE'12.

3 Mahmud, Junayed, et al. "On Using GUI Interaction Data to Improve Text Retrieval-based Bug Localization." ICSE 24.

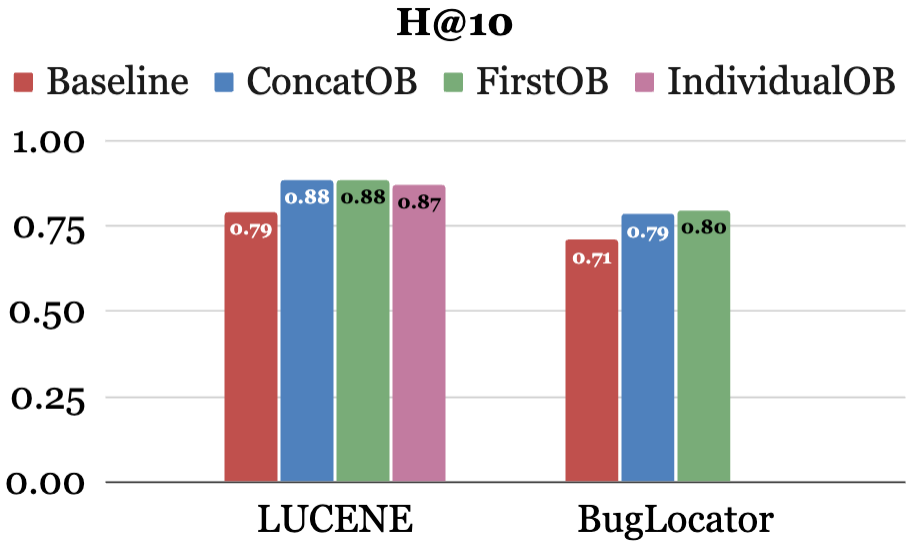
RQ4: Buggy Code Localization Results



RQ4: Buggy Code Localization Results



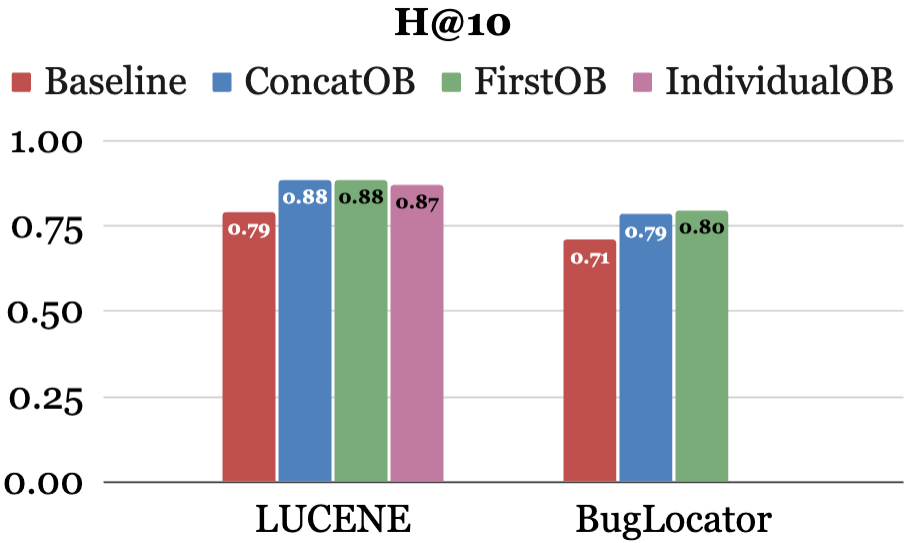
RQ4: Buggy Code Localization Results



RQ4: Buggy Code Localization Results

Best Pipelines

LUCENE: ConcatOB & FirstOB
BugLocator: FirstOB



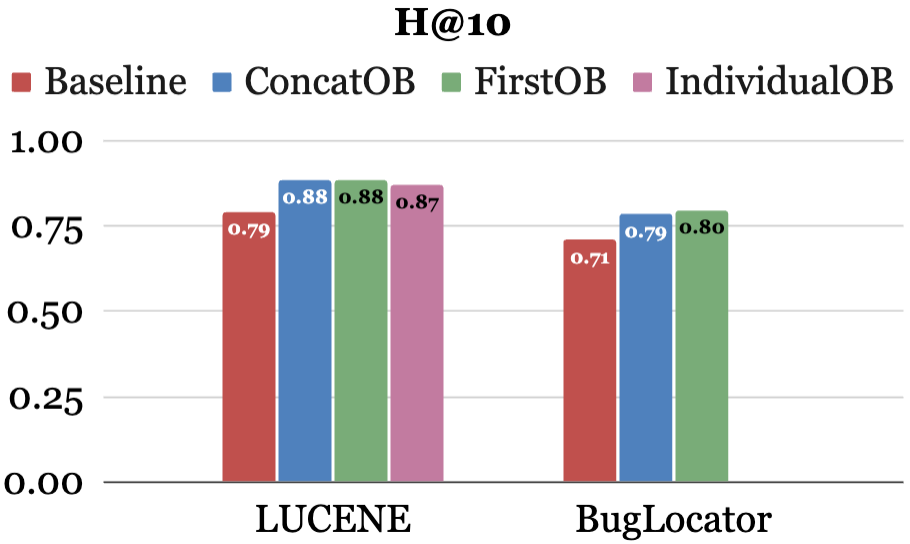
RQ4: Buggy Code Localization Results

Best Pipelines

LUCENE: ConcatOB & FirstOB
BugLocator: FirstOB

Outperforms Baselines

LUCENE: By 11.49% (H@10)
BugLocator: 12.41% (H@10)



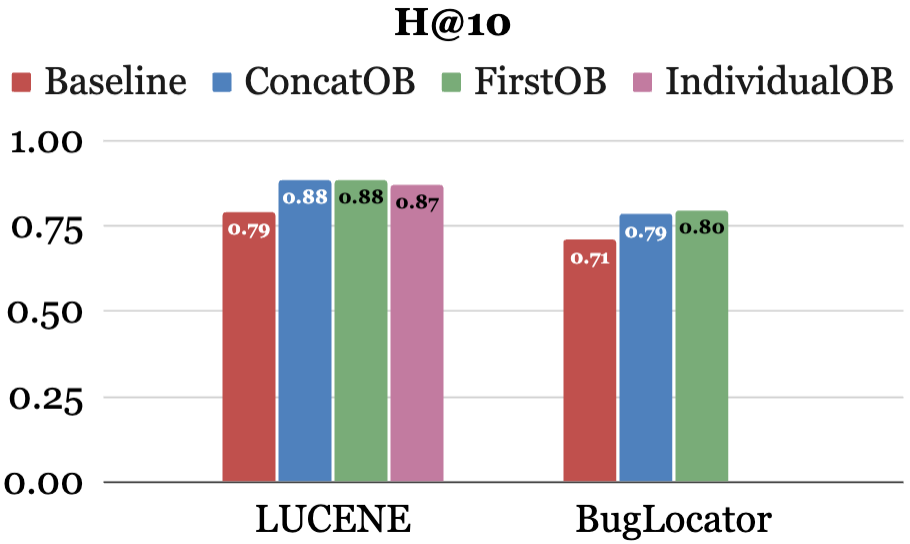
RQ4: Buggy Code Localization Results

Best Pipelines

LUCENE: ConcatOB & FirstOB
BugLocator: FirstOB

Outperforms Baselines

LUCENE: By 11.49% (H@10)
BugLocator: 12.41% (H@10)



Buggy UI Localization can automate & improve UI-based Buggy Code Localization

Summary



Summary

Buggy UI Localizatio n (S CL)



Summary

Buggy UI Localization (S[illegible] CL)



Constructed a Novel Dataset



Summary

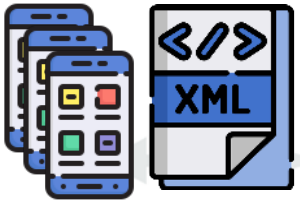
Buggy UI Localizatio n (S… CL)



Constructe d a Novel Dataset



Investigate d Information Sources



Summary

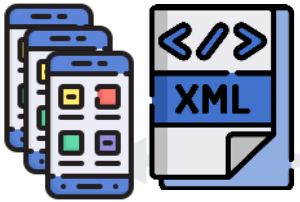
Buggy UI Localization (S[illegible] CL)



Constructed a Novel Dataset



Investigated Information Sources



Bug Localization Using SL

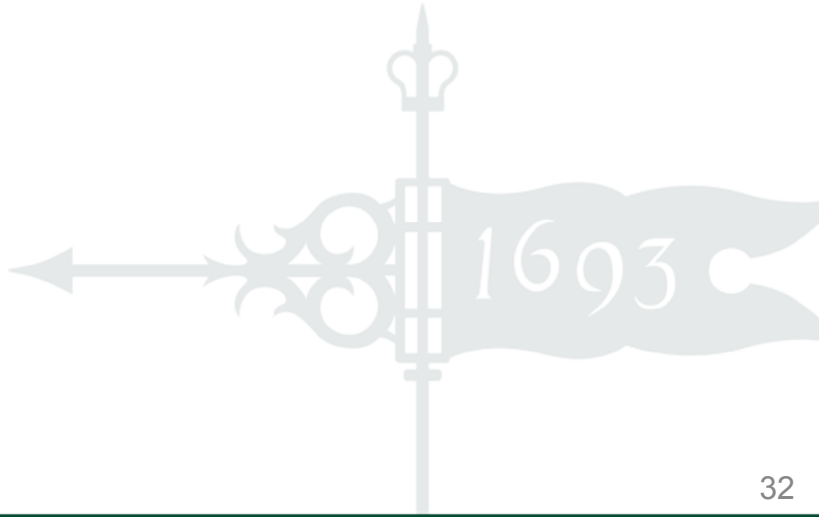


Take Aways



Take Aways

Existing DL approaches can be used for Buggy UI Localization tasks (SL & CL)



Take Aways

Existing DL approaches can be used for Buggy UI Localization tasks (SL & CL)

Both visual & textual sources of information are effective for Buggy UI Localization

Take Aways

Existing DL approaches can be used for Buggy UI Localization tasks (SL & CL)

Both visual & textual sources of information are effective for Buggy UI Localization

Buggy UI Localization can automate & improve Bug Localization



SCAN ME

Let's get connected!

Antu Saha
asaha02@wm.edu