

Spatially Combined Keyword Searches

Artur Titkov

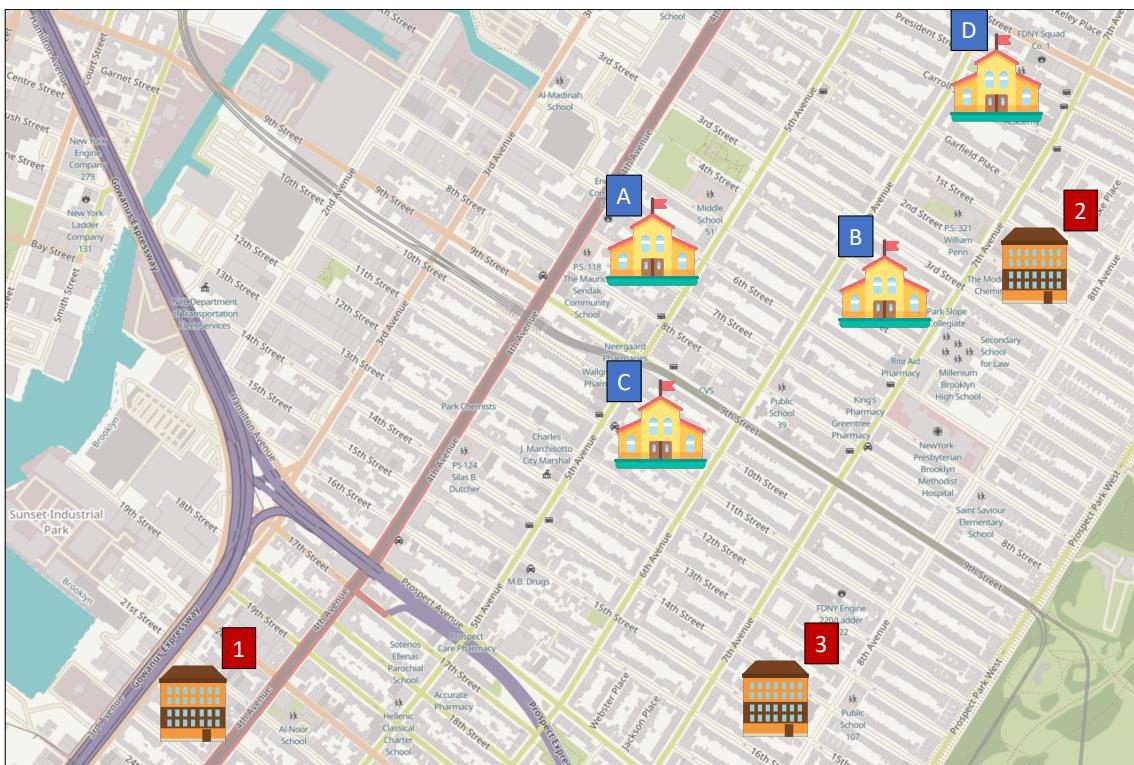
Panagiotis Bouros

JOHANNES GUTENBERG
UNIVERSITÄT MAINZ



Motivation

apartments	amenities
1	balcony, pool, gym
2	balcony, gym
3	balcony, pool



kindergartens	description
A	German, indoor-playground
B	German, outdoor-playground
C	German, outdoor-playground
D	outdoor-playground

$$q = \{\Psi_{\text{apartments}}, \Psi_{\text{kindergartens}}, \epsilon\}$$

$$\Psi_{\text{apartments}} = \{\text{balcony, gym}\}$$

$$\Psi_{\text{kindergartens}} = \{\text{german, outdoor-playground}\}$$

$$\epsilon = 1\text{km}$$

$$\text{dist}(1, A) = 1.8\text{km}$$

$$\text{dist}(1, C) = 1.2\text{km}$$

$$\text{dist}(2, B) = 300\text{m}$$

$$\text{dist}(2, C) = 900\text{m}$$

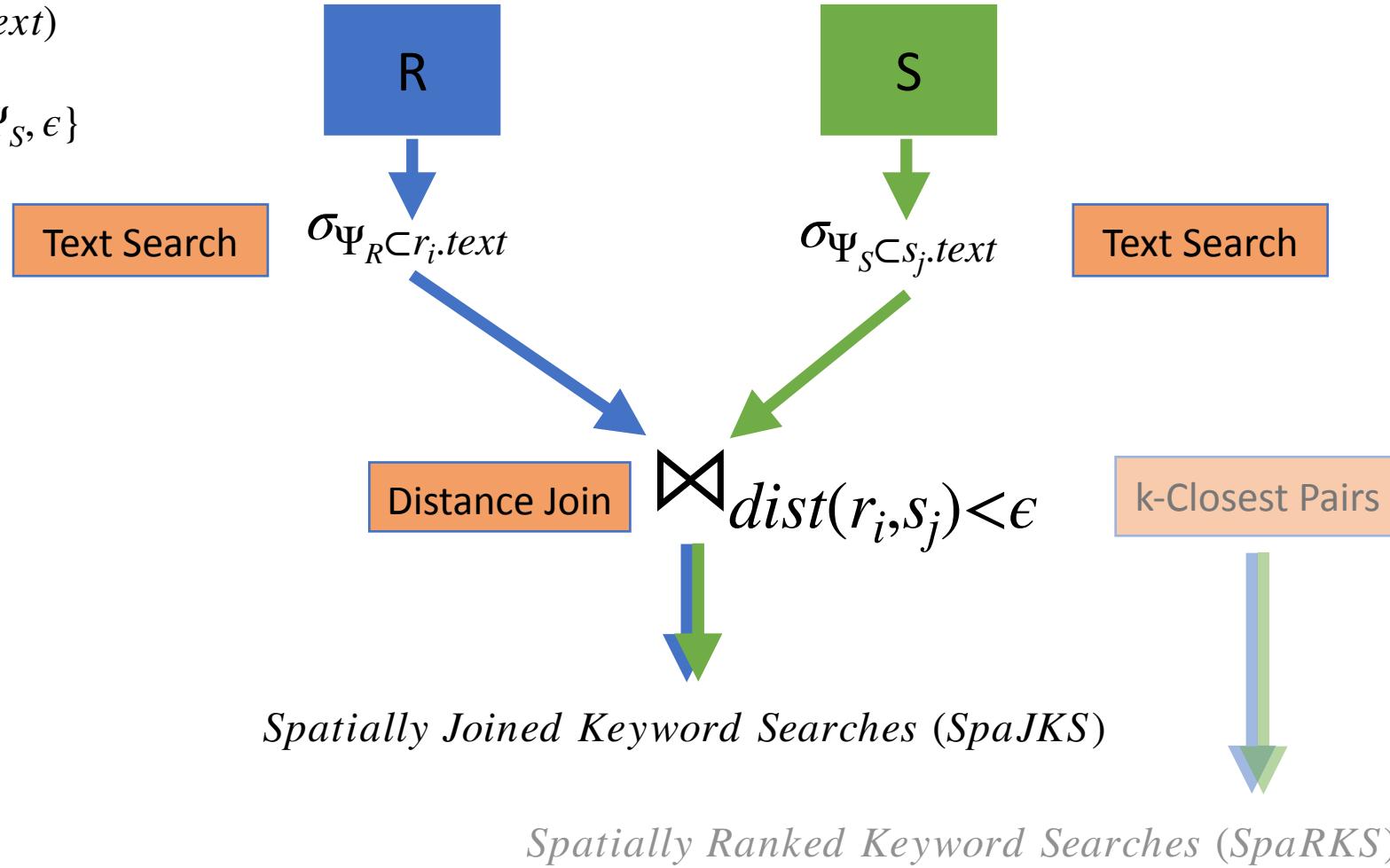
-
- Problem definition
 - Evaluation strategies
 - Possible Indices
 - Text/Spatial First
 - Spatial/Spatio-textual Probing
 - Join Based
 - Experimental results

Problem Definition

JG|U

$$r_i = (x, y, \text{text})$$

$$q = \{\Psi_R, \Psi_S, \epsilon\}$$



Evaluation Strategies



Inverted File | R-Tree | IR-Tree[1] | Spatio-textual Grid [2]

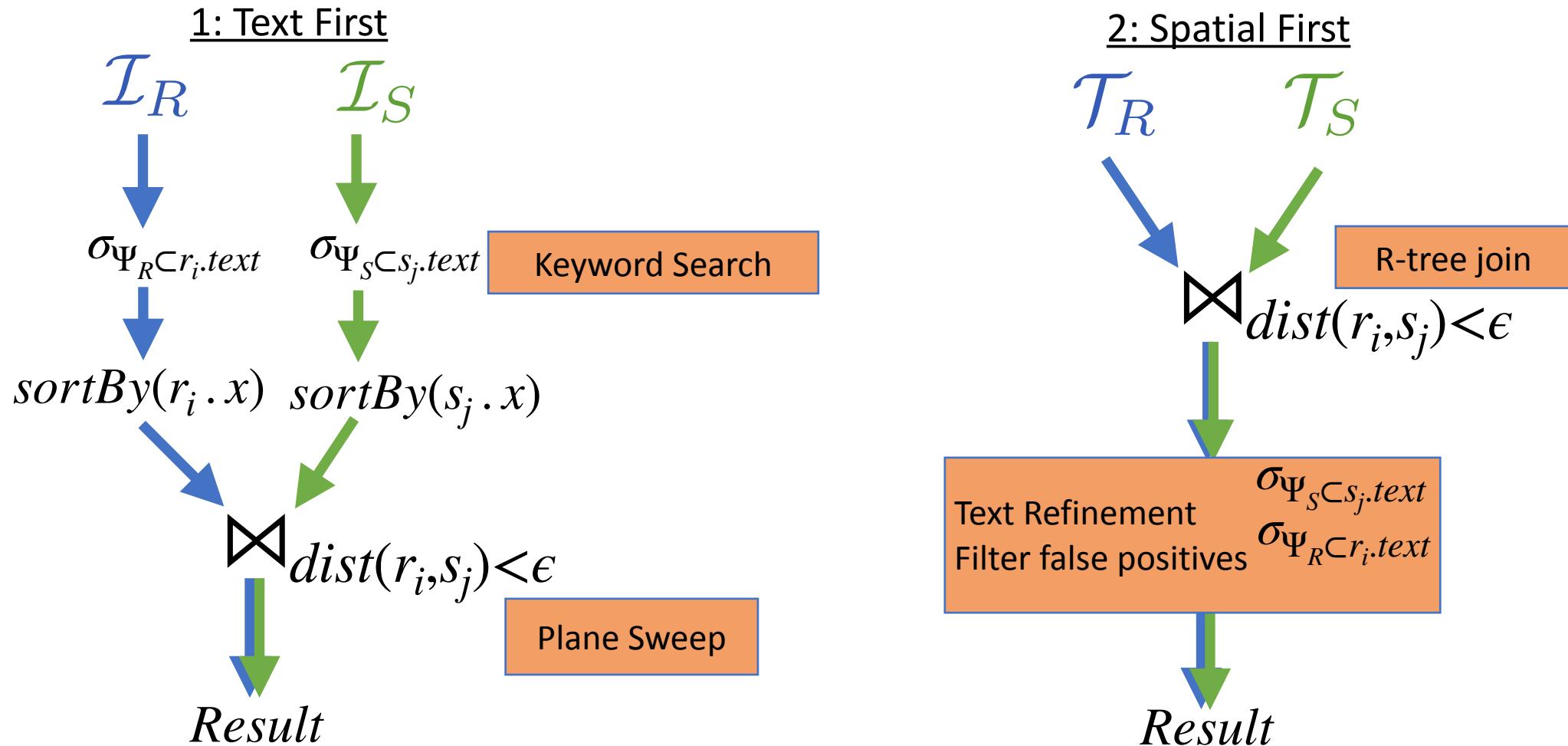
\mathcal{I}	\mathcal{T}	\mathcal{H}	G
\mathcal{I}_R	1: Text first		\mathcal{I}_S
\mathcal{T}_R	2: Spatial first		\mathcal{T}_S
\mathcal{I}_R	3: Spatial probing		\mathcal{T}_S
\mathcal{I}_R	4: Spatio-textual probing		\mathcal{H}_S
\mathcal{H}_R	5: IR-tree join		\mathcal{H}_S
G_R	6: Grid based join		G_S

[1] G. Cong, C. S. Jensen, and D. Wu. Efficient Retrieval of the Top-k Most Relevant Spatial Web Objects. Proc. VLDB Endow. 2, 1 (2009), 337–348.

[2] P. Bouros, S. Ge, and N. Mamoulis. Spatio-textual similarity joins. Proc. VLDB Endow. 6, 1 (2012), 1–12.

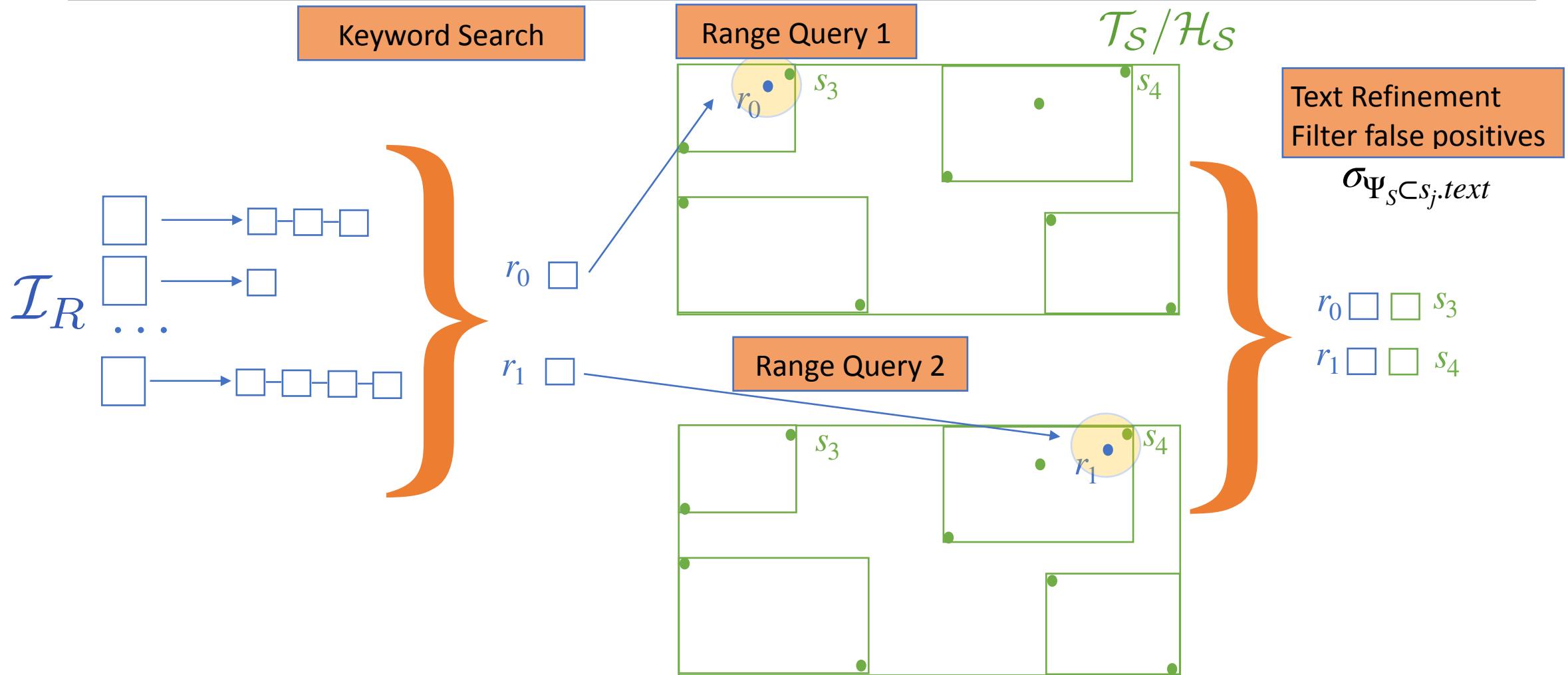
1/2: Text/Spatial First

JG|U



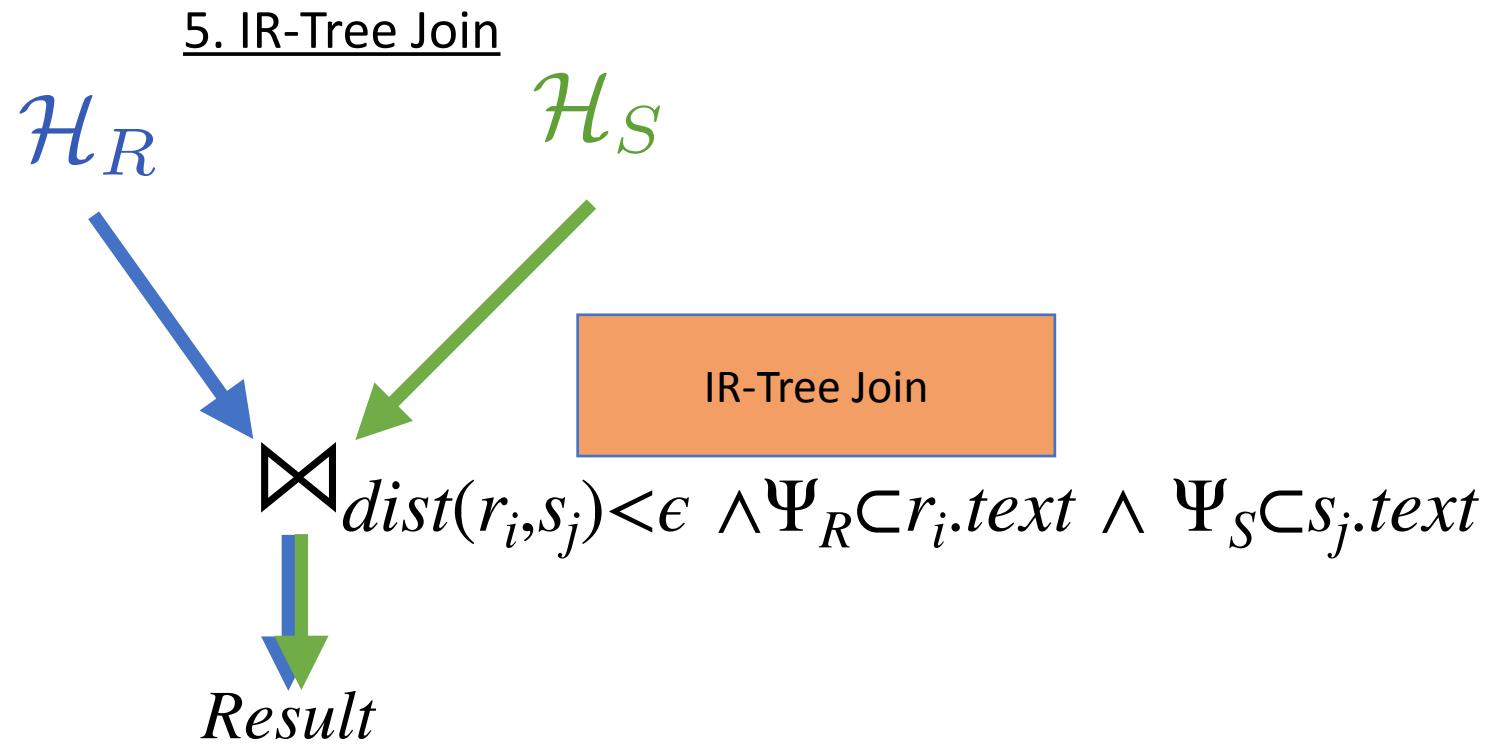
3/4: Spatial/Spatio-textual Probing

JG|U

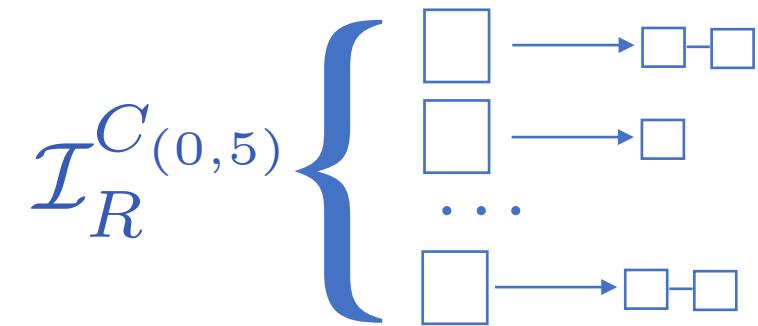
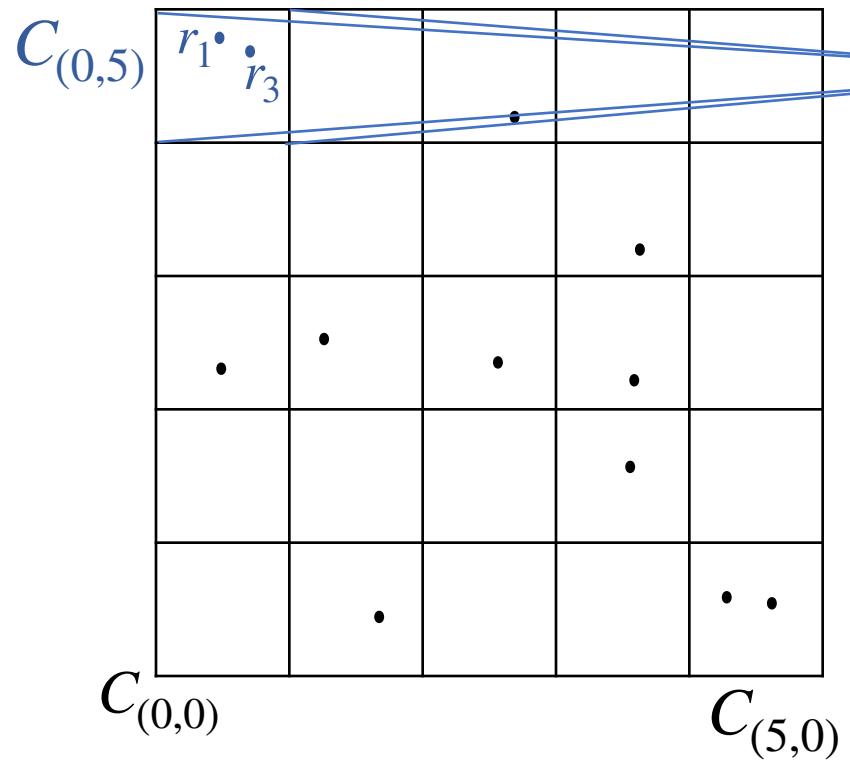


5: IR-Tree Join

JG|U

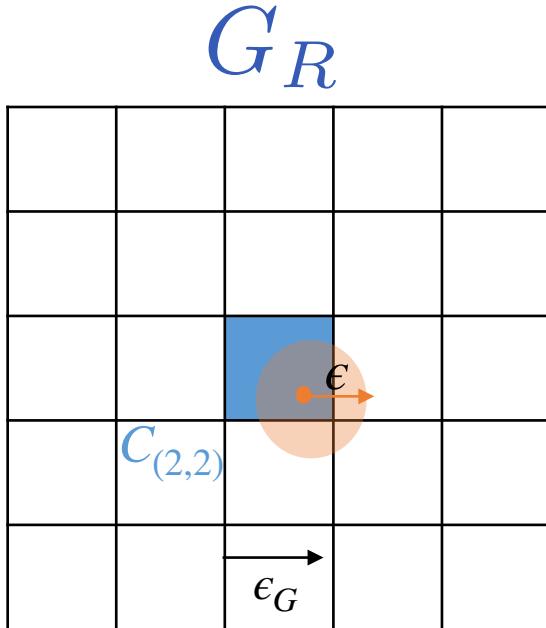


6: Spatio-Textual Grid

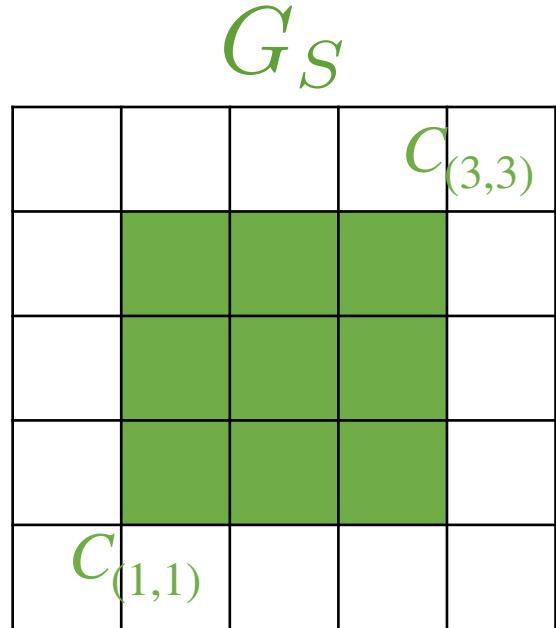


6: Grid-based Join

JG|U



$$\epsilon < \epsilon_G$$



- iterate over cells R
- identify join partner of S
- per pair -> test first

$$G_R^{C_{(3,3)}} \bowtie_{dist(r_i, s_j) < \epsilon} G_R^{C_{(1,1)}}$$

...

$$G_R^{C_{(3,3)}} \bowtie_{dist(r_i, s_j) < \epsilon} G_R^{C_{(3,3)}}$$

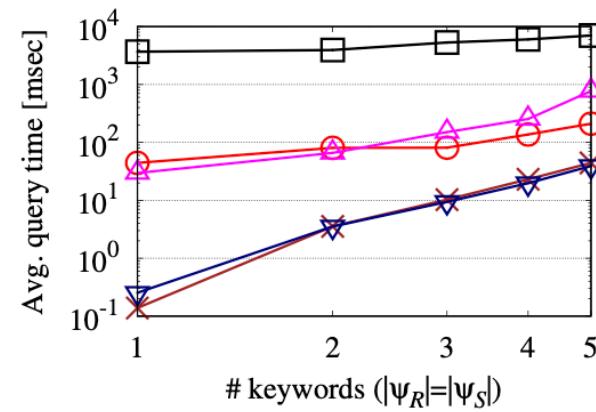
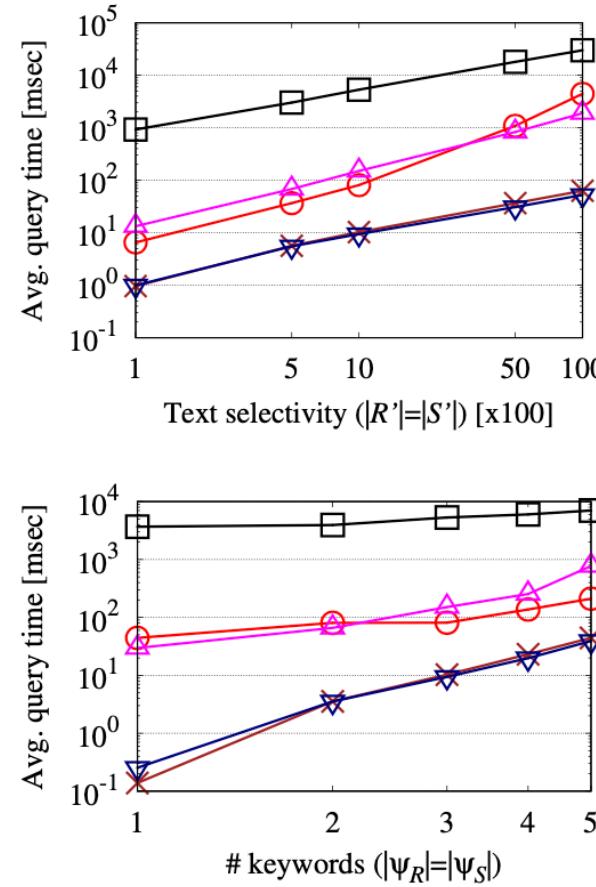
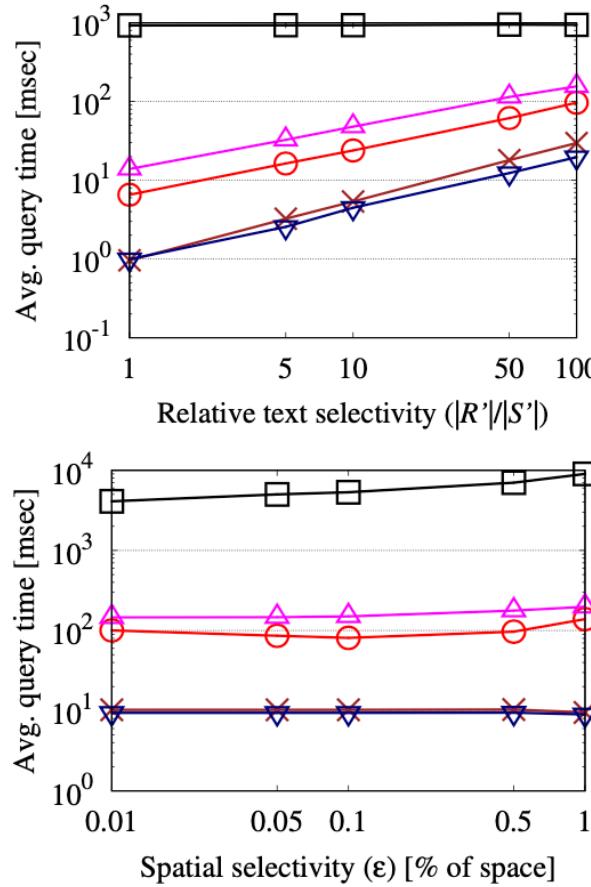
Experiments



- Datasets
 - one real (polybot)[3]
 - multiple synthetic
- AVG response time when varying
 - spatial selectivity
 - text selectivity
 - number keywords
 - cardinality
 - ...

[3] M. Christoforaki, . He, C. Dimopoulos, A. Markowetz, and T. Suel. Text vs. space: efficient geo-search query processing. ACM CIKM 2009. 423–432.

Experiments Polybot



3 tiers on response time:

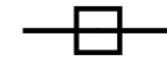
1. text first / grid join



2. spatio-textual probing/ IR-tree join



3. spatial-probing / spatial first



To sum up

JG|U

-
- Contributions
 - new retrieval tasks
 - Spatially Joined Keyword Searches (SpaJKS)
 - Spatially Ranked Keyword Searches (SpaRKS)
 - Future work
 - multiple keyword searches
 - text search as similarity search