

Feedforward-Aided Course Designs for Similarity Search

Thomas Hütter and Daniel Kocher

University of Salzburg, Austria

2nd Int. Workshop on Data Systems Education (DataEd'23)

Seattle, WA, USA

June 23, 2023

Course Facts:

- *Similarity Search in Large Databases*
- 2.5 ECTS (\approx 62.5 hours), 20–30 Master's level students per year
- Distance measures, lower/upper bounds, and similarity indexes

¹

Bayardo et al. *Scaling up all pairs similarity search*. WWW 2007. <https://doi.org/10.1145/1242572.1242591>

Course Facts:

- *Similarity Search in Large Databases*
- 2.5 ECTS (\approx 62.5 hours), 20–30 Master's level students per year
- Distance measures, lower/upper bounds, and similarity indexes

Learning by Design ■ Functional Artifact ■ Set Similarity Joins¹

¹ Bayardo et al. *Scaling up all pairs similarity search*. WWW 2007. <https://doi.org/10.1145/1242572.1242591>

Set Similarity Joins

Given two collections R and S , a distance threshold ϵ under a function $d(.,.)$:

$$\{(r, s) \in R \times S \mid d(r, s) \leq \epsilon\}$$

R

r_1	1	3	5		
r_2	1	2	3	4	
r_3	1	2	4	9	11

Jaccard $d_j = \frac{|r \cap s|}{|r \cup s|}$

$\epsilon_j = 0.5$



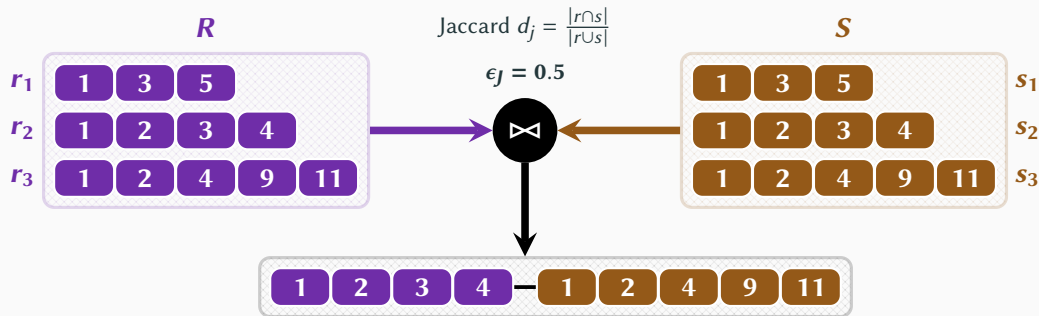
S

	1	3	5			s_1
	1	2	3	4		s_2
	1	2	4	9	11	s_3

Set Similarity Joins

Given two collections R and S , a distance threshold ϵ under a function $d(.,.)$:

$$\{(r, s) \in R \times S \mid d(r, s) \leq \epsilon\}$$



Course Design

**Project-Based
Learning (PBL)**

**Task-Based
Learning (TBL)**

Course Design

**Project-Based
Learning (PBL)**



**Task-Based
Learning (TBL)**

Course Design

Project-Based Learning (PBL)



Task-Based Learning (TBL)



Project-Based Learning (PBL)



Task-Based Learning (TBL)



PBL vs. TBL – The Case for AllPairs²

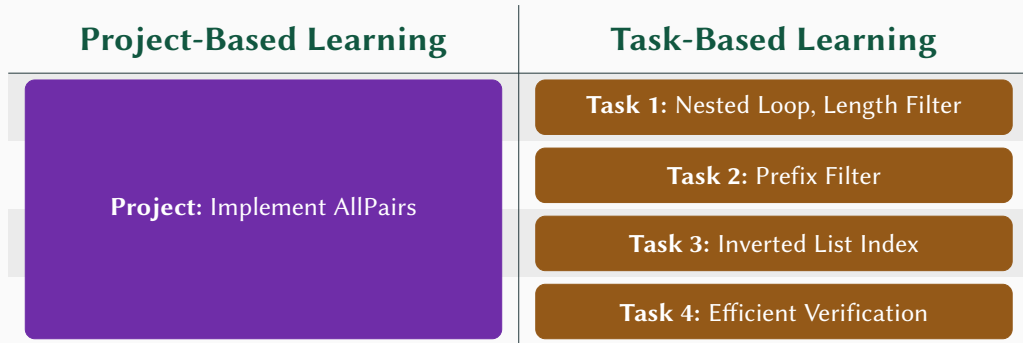


Figure 1: Implementation of the AllPairs² algorithm – PBL vs. TBL.

²

Bayardo et al. *Scaling up all pairs similarity search*. WWW 2007. <https://doi.org/10.1145/1242572.1242591>

Experiences



Auto-Grader System



Continuous and immediate **feedforward**.



Automated basis for **grading**.



Support for **heterogeneity**.



Motivation by (unexpected) competition.

Auto-Grader System



Continuous and immediate **feedforward**.



Automated basis for **grading**.



Support for **heterogeneity**.



Motivation by (unexpected) competition.

Take Away: An auto-grader system is indispensable.

Heterogeneous Groups



Accounting for **different backgrounds** is **challenging**.



Programming knowledge vs. conceptualization.



Students may **complement** each other's **strengths**.

Heterogeneous Groups



Accounting for **different backgrounds** is **challenging**.



Programming knowledge vs. conceptualization.



Students may **complement** each other's **strengths**.

Take Away: Heterogeneity is an opportunity.

Individualization & Group Size



Individual feedforward is good but **time-consuming**.



Scaling individual feedforward **to large groups** is **hard**.

Individualization & Group Size



Individual feedforward is good but **time-consuming**.



Scaling individual feedforward **to large groups** is **hard**.

Future Prospect: Extended diagnosis capabilities for auto-grader.

Student Evaluation

Various criteria: Degree of difficulty, relevance of topic, teaching material quality, ...

Cohorts: 10–27 students

Scale: 1–7 (higher is better)

Student Evaluation

Various criteria: Degree of difficulty, relevance of topic, teaching material quality, ...

Cohorts: 10–27 students

Scale: 1–7 (higher is better)

Highlights

	Relevance	Goal	Overall	Support	Objectives	Material
PBL		↑ 1.4%		94%	85%	89%
TBL	↑ 8.3%		↑ 5.9%			

Conclusion

Two **feedforward-aided course designs** for **similarity search**.

Project-Based vs. Task-Based Learning.

Experiences for **both designs** in class.

Both designs are **suitable** for teaching **similarity search**.

Auto-grader and active **communication channels** are indispensable.

Feedforward-Aided Course Designs for Similarity Search

Questions?

✉ thomas.huetter@plus.ac.at

✉ dkocher@cs.sbg.ac.at

