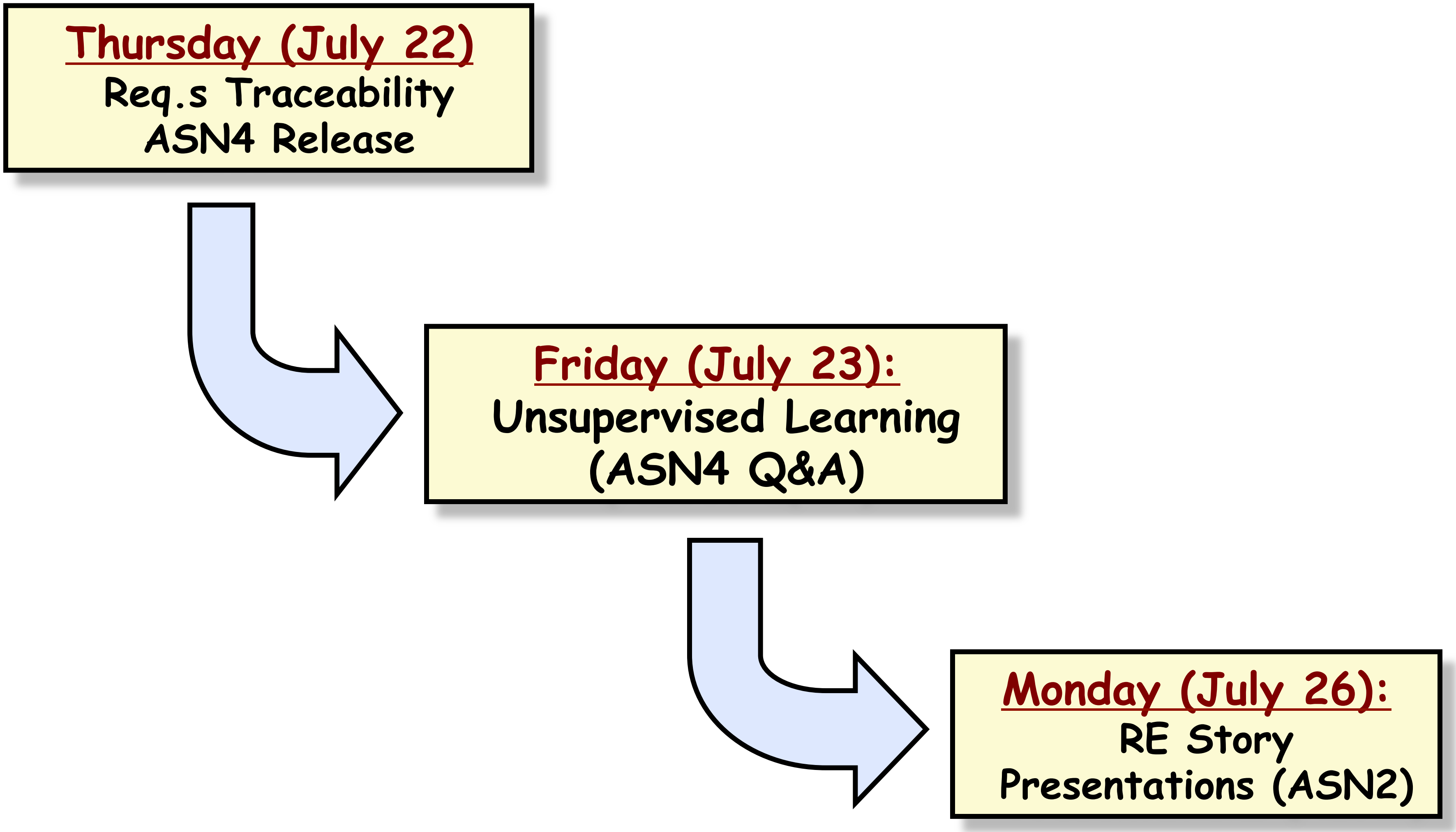
Requirements Engineering

(Summer 2021)

Prof. Nan Niu (nan.niu@uc.edu)

https://github.com/nanniu/RE-Summer2021

Today's Menu

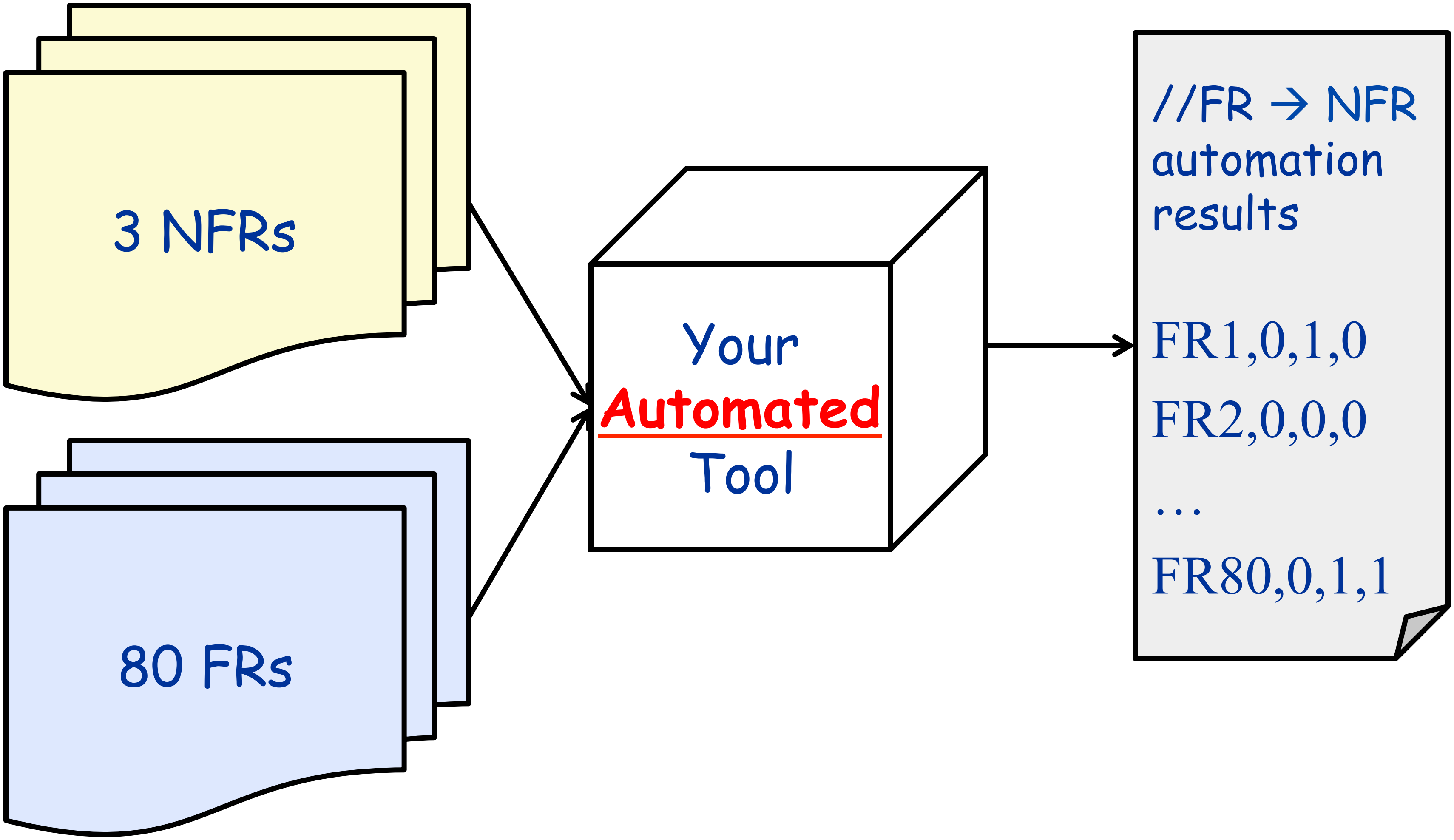


Thursday (July 22)
Req.s Traceability
ASN4 Release

Friday (July 23):
Unsupervised Learning
(ASN4 Q&A)

Monday (July 26):
RE Story
Presentations (ASN2)

ASN4: A conceptual picture



3 NFRs

80 FRs

Your **Automated** Tool

//FR → NFR automation results

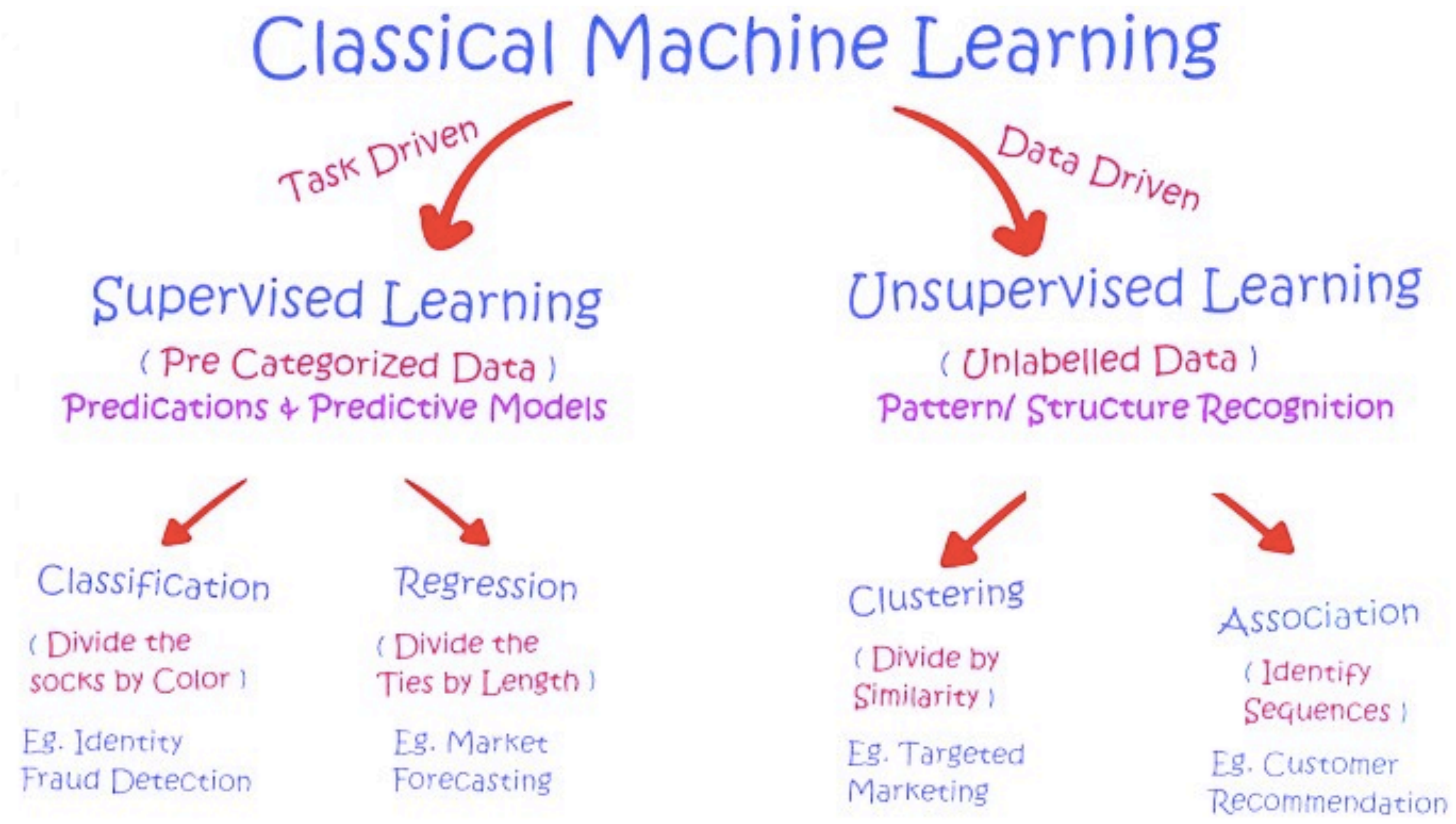
FR1,0,1,0

FR2,0,0,0

…

FR80,0,1,1

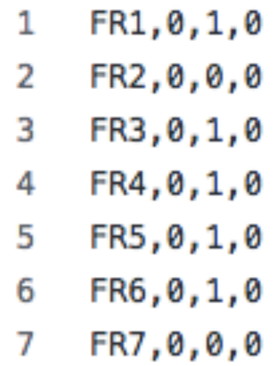
ML: Will it work for ASN4?



Source: *Google Images*

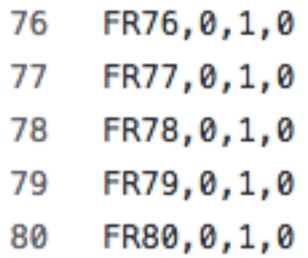
Supervised Learning: *not really*

- Run #1: 80 FRs and 3 NFRs
- Run #2: 100 FRs and 3 NFRs (i.e., 20 new/unseen FRs compared to Run #1)
- Run #3: 100 FRs and 4 NFRs (i.e., 1 new/unseen NFR compared to Run #2)



```
FR1,0,1,0
FR2,0,0,0
FR3,0,1,0
FR4,0,1,0
FR5,0,1,0
FR6,0,1,0
FR7,0,0,0
```

...

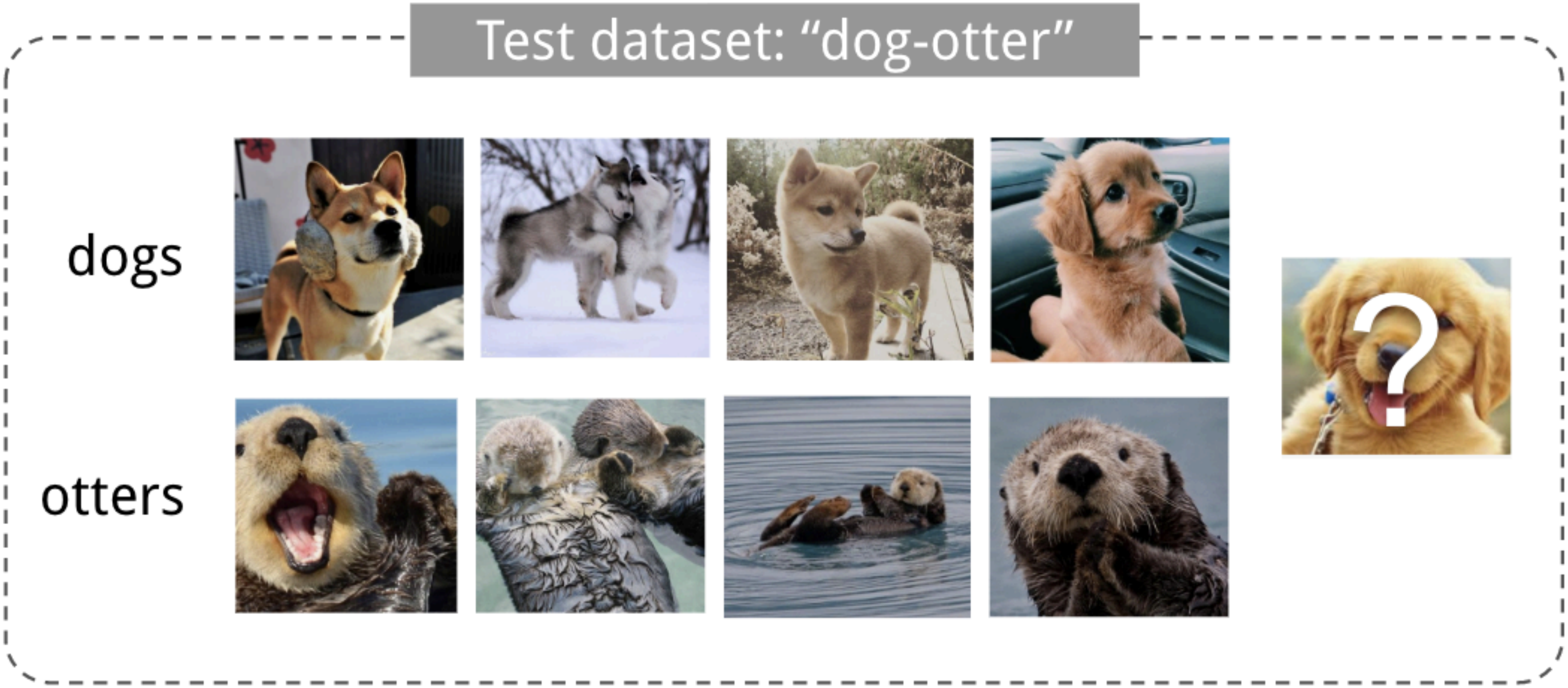


```
FR76,0,1,0
FR77,0,1,0
FR78,0,1,0
FR79,0,1,0
FR80,0,1,0
```

(unseen data)

FR81: “description of a new FR”

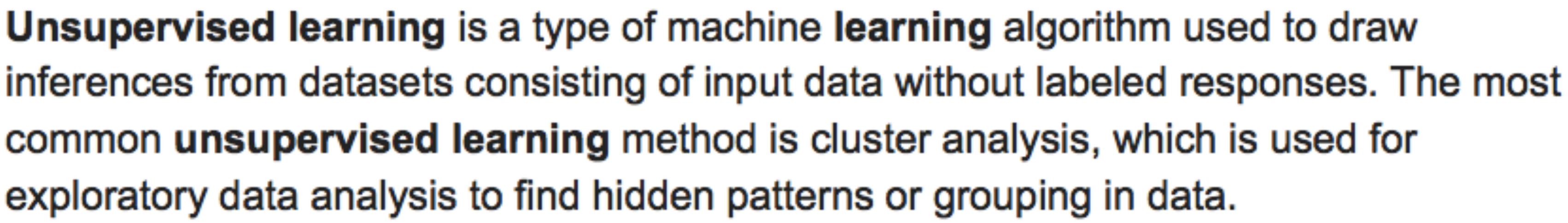
What to learn?



Source: *https://lilianweng.github.io/lil-log/2018/11/30/meta-learning.html*

Today's Take-Away

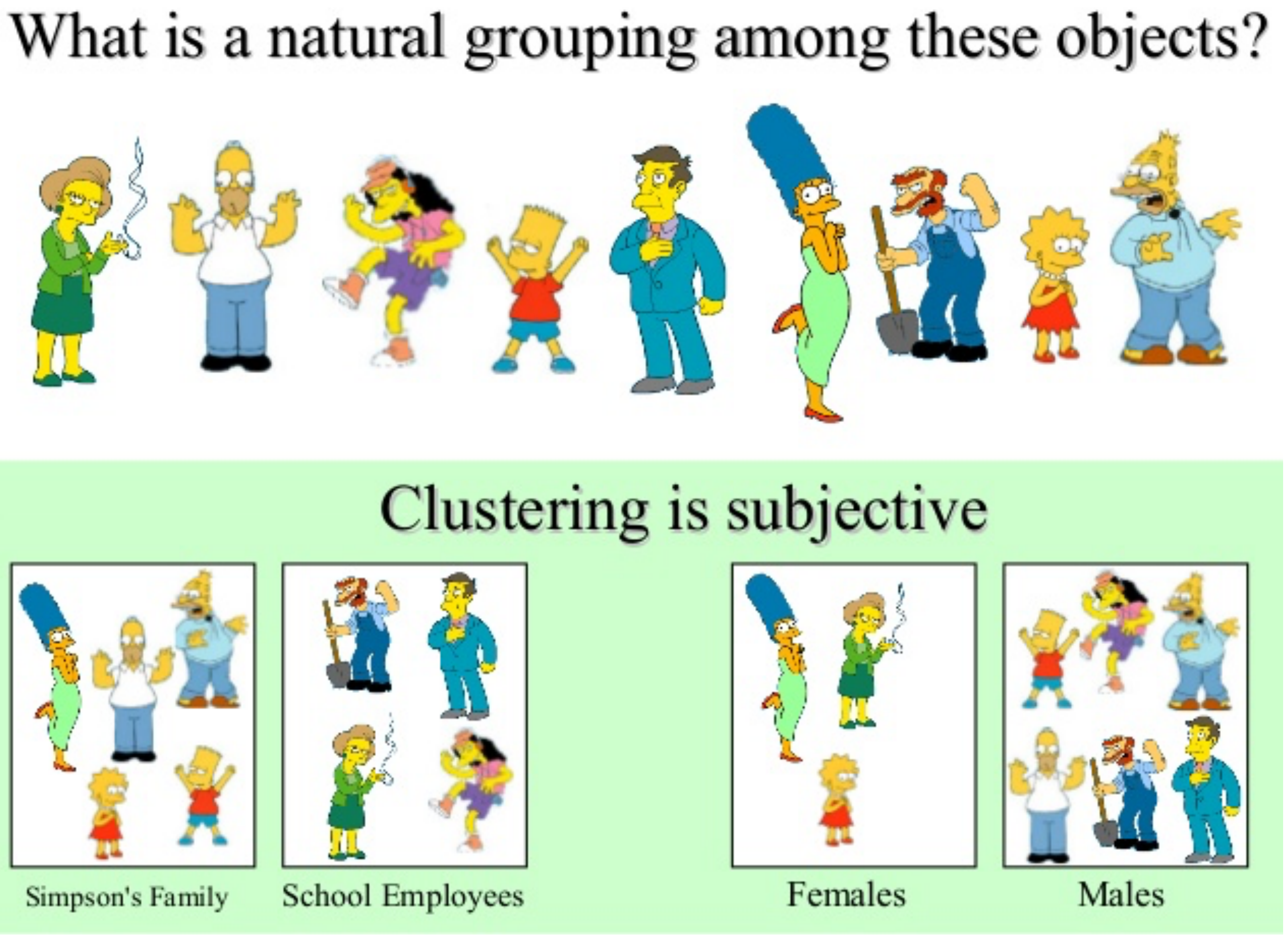
- →Unsupervised learning can be used to solve ASN4.



Unsupervised learning is a type of machine **learning** algorithm used to draw inferences from datasets consisting of input data without labeled responses. The most common **unsupervised learning** method is cluster analysis, which is used for exploratory data analysis to find hidden patterns or grouping in data.

Source: *Google Search*

Clustering: Finding ~~the~~ a natural grouping of data

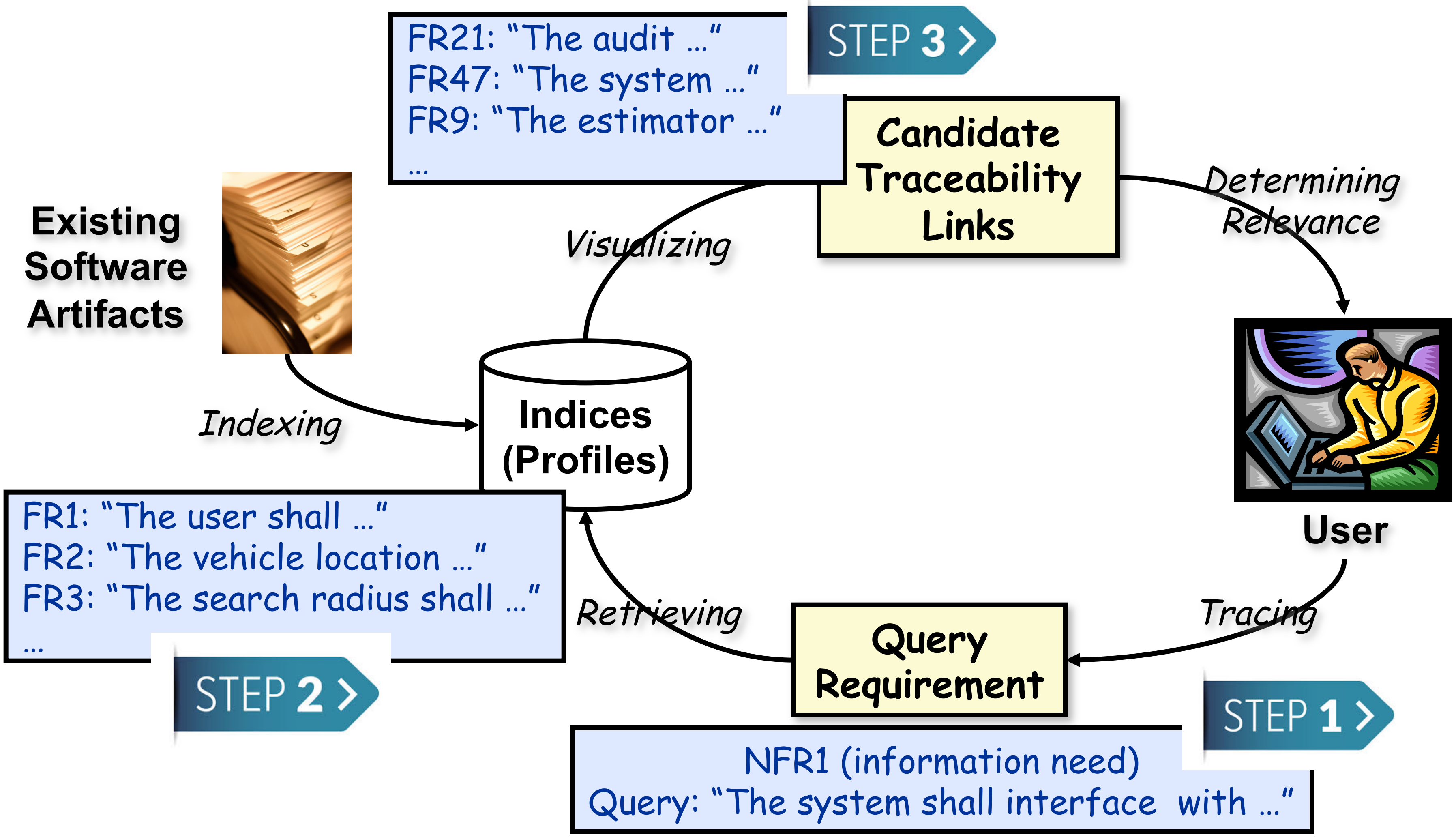


What is a natural grouping among these objects?

Clustering is subjective

Source: *Google Images*

IR-Based ASN4 Solution (fully automatic)



Existing Software Artifacts

Indexing

Indices (Profiles)

FR1: "The user shall …"
FR2: "The vehicle location …"
FR3: "The search radius shall …"
…

STEP 2

Retrieving

Query Requirement

Tracing

User

STEP 1

NFR1 (information need)
Query: "The system shall interface with …"

Visualizing

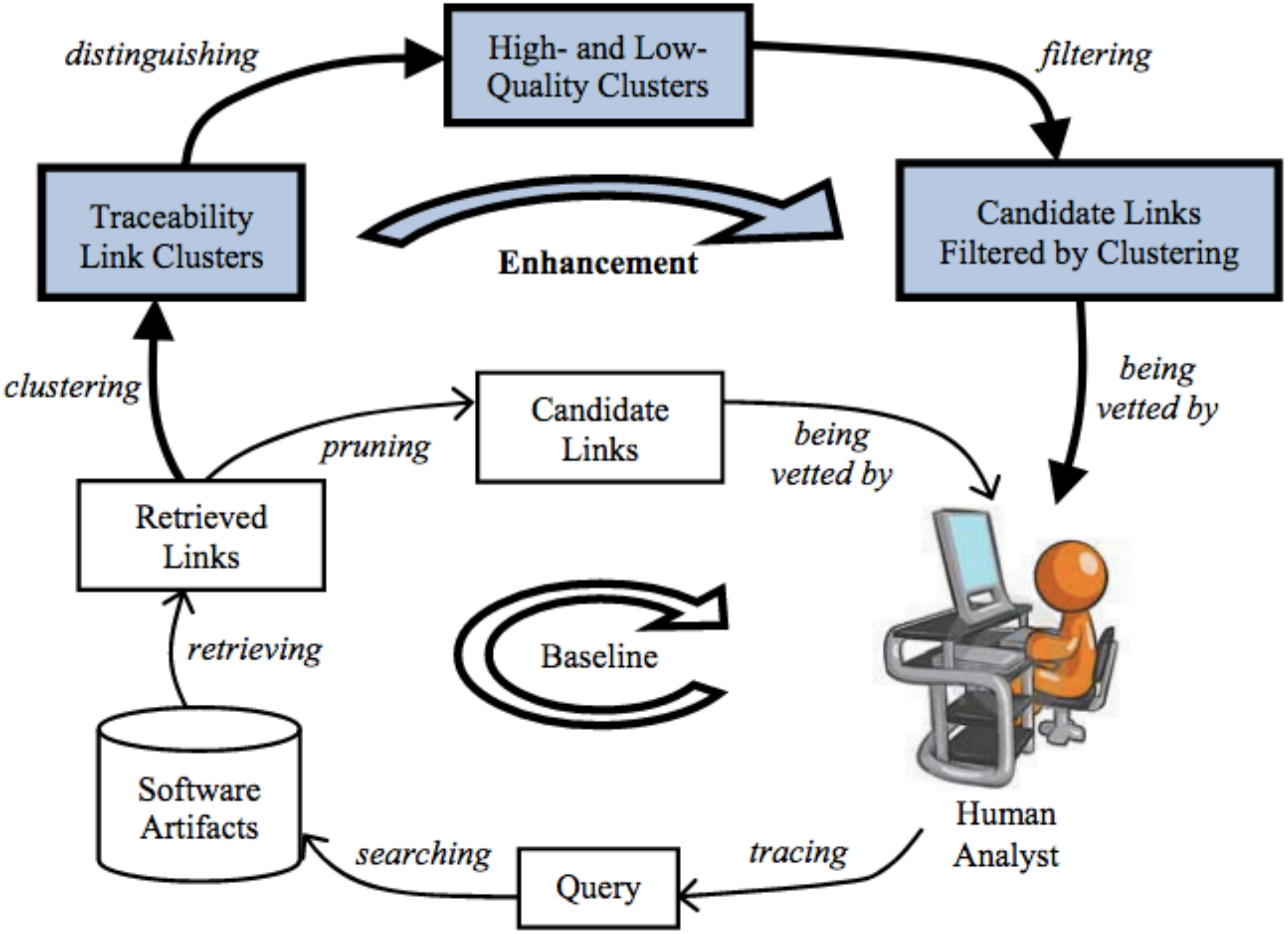
Candidate Traceability Links

Determining Relevance

FR21: "The audit …"
FR47: "The system …"
FR9: "The estimator …"
…

STEP 3

Clustering as Enhancement



Source: https://homepages.uc.edu/~niunn/papers/RE12.pdf

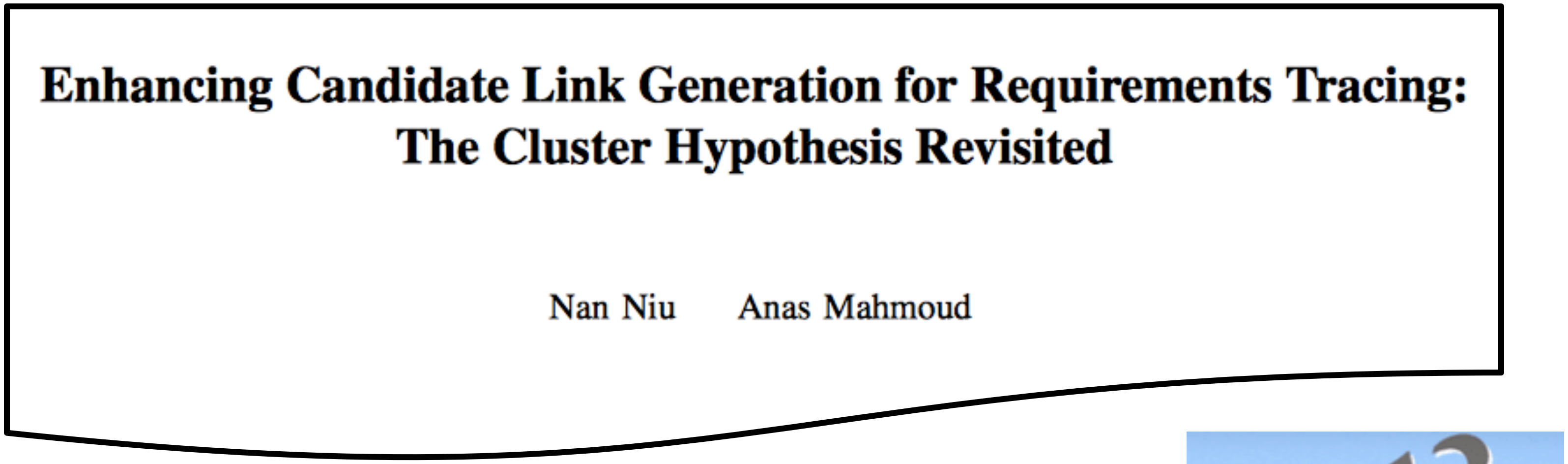
Cluster Hypothesis

- In IR, *cluster hypothesis* suggests, "relevant documents tend to cluster near other relevant documents and farther away from irrelevant ones".
  - Applying *cluster hypothesis* in automated requirements tracing, it suggests, "correct links tend to be more similar to each other than to incorrect links".

Source: https://homepages.uc.edu/~niunn/papers/RE12.pdf

Key questions answered

Does cluster hypothesis hold in automated requirements traceability, and if so, how to best exploit it?



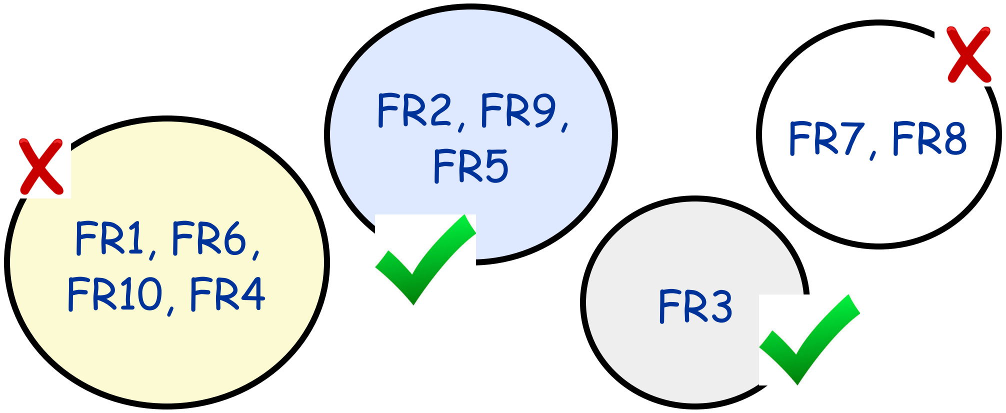
Enhancing Candidate Link Generation for Requirements Tracing: The Cluster Hypothesis Revisited

Nan Niu Anas Mahmoud

Source: https://homepages.uc.edu/~niunn/papers/RE12.pdf

Conceptually

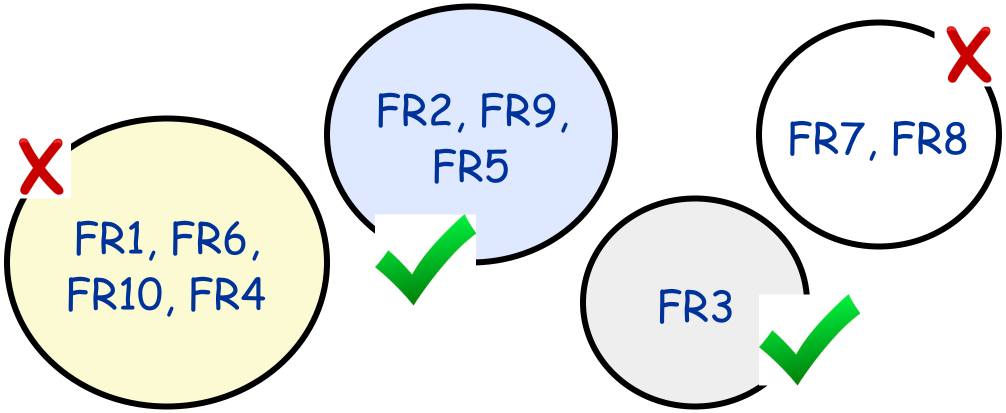
For a particular trace query (e.g., an NFR), identify candidate traceability links (e.g., via Jaccard index). Then, cluster the candidate traceability links:



each cluster (as opposed to each link / FR) will be judged "correct / traceable" or "incorrect / not traceable".

Today's poll question: ***"In addition to which clustering algorithm to use, <u>how many</u> other decisions are required?"***

Conceptually, to achieve *full automation*



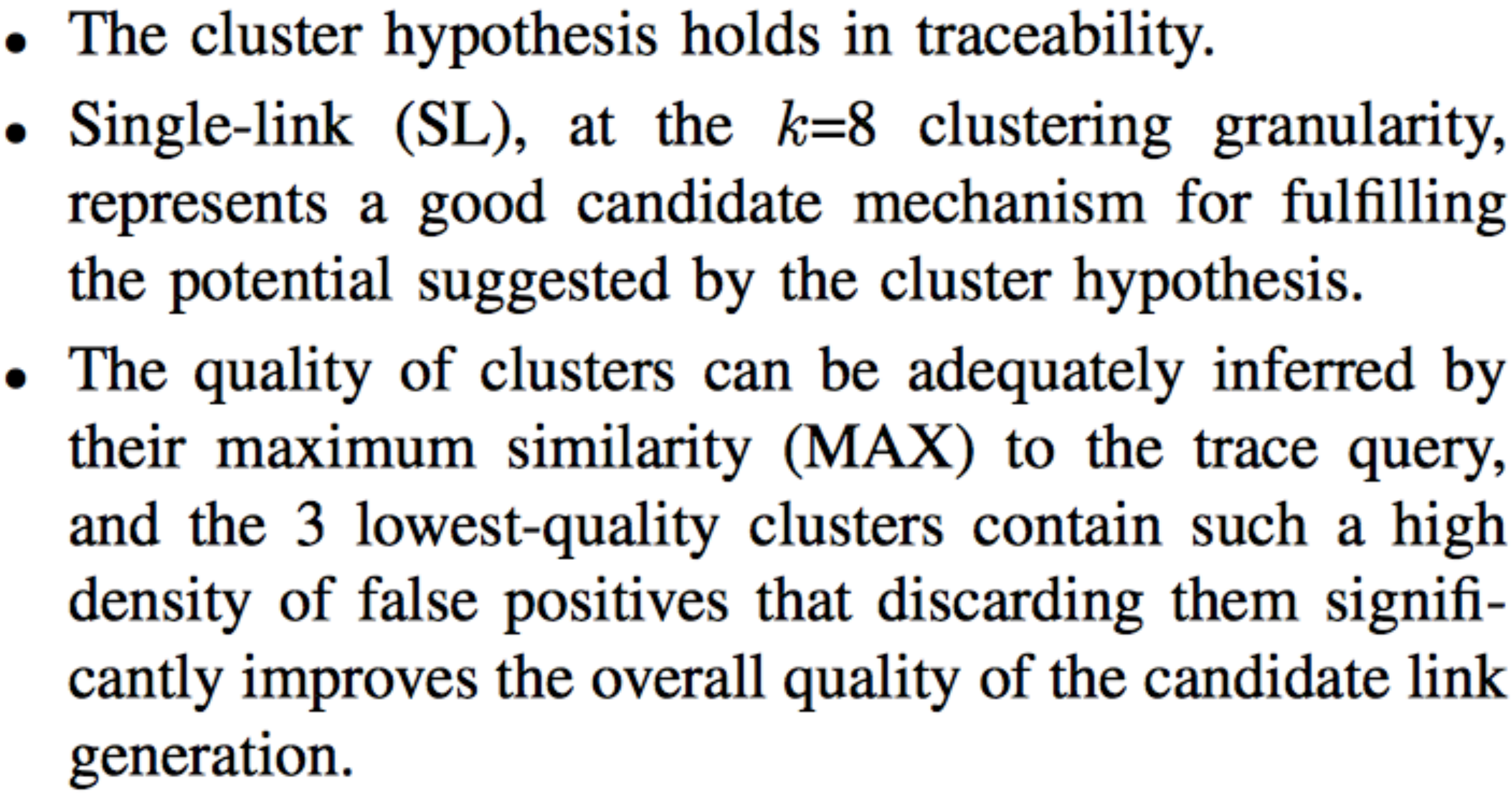
(0) Which clustering algorithm to use?

(1) How many clusters (*k*) to produce?

(2) How to rank the resulting clusters?

(3) After ranking, how many "good" clusters to keep and how many "bad" clusters to remove?

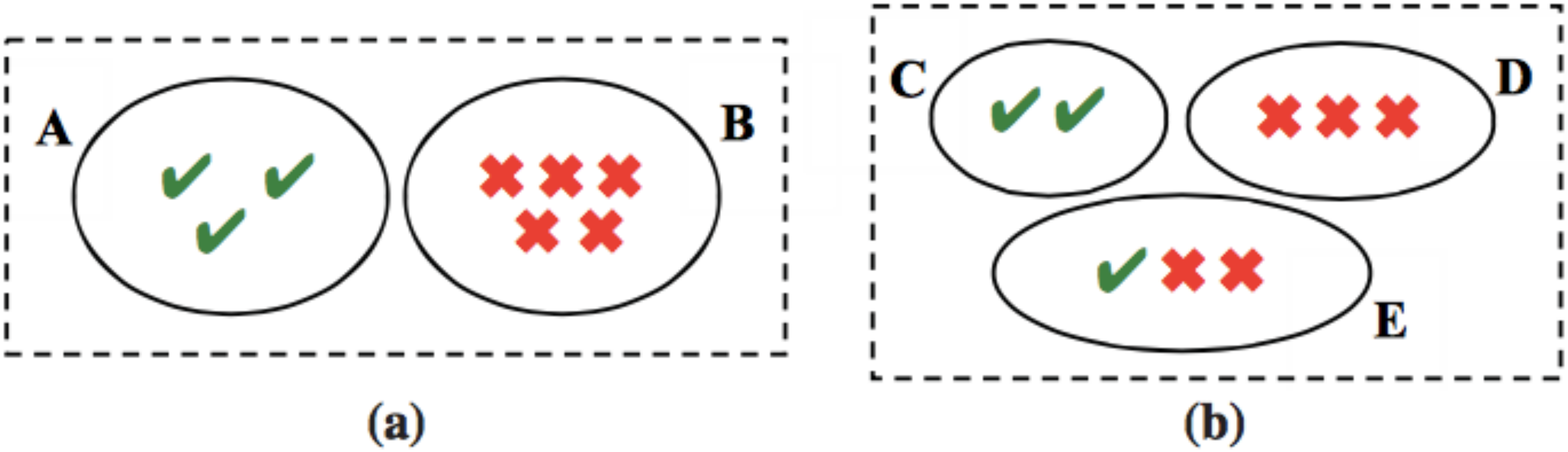
Results



- The cluster hypothesis holds in traceability.
- Single-link (SL), at the k=8 clustering granularity, represents a good candidate mechanism for fulfilling the potential suggested by the cluster hypothesis.
- The quality of clusters can be adequately inferred by their maximum similarity (MAX) to the trace query, and the 3 lowest-quality clusters contain such a high density of false positives that discarding them significantly improves the overall quality of the candidate link generation.

Source: *https://homepages.uc.edu/~niunn/papers/RE12.pdf*

Evaluating Clustering Results



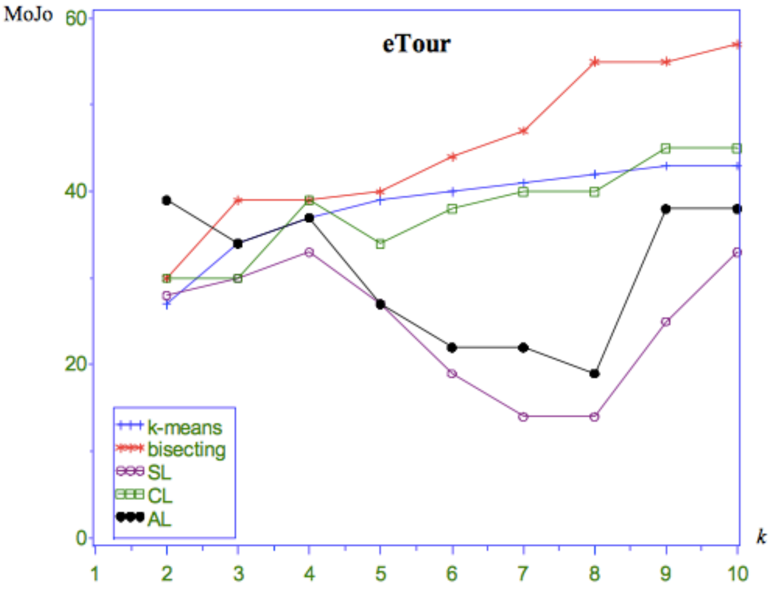
Suppose (a) is the "answer set", then the quality of (b) can be measured by the MoJo distance (i.e., the number of Moves and Joins to transform (b) to (a)); here MoJo distance = 2 (i.e., move the correct link from E to C, and then join D and the revised E together).

The smaller the MoJo distance, the better the clustering.

Source: *https://homepages.uc.edu/~niunn/papers/RE12.pdf*

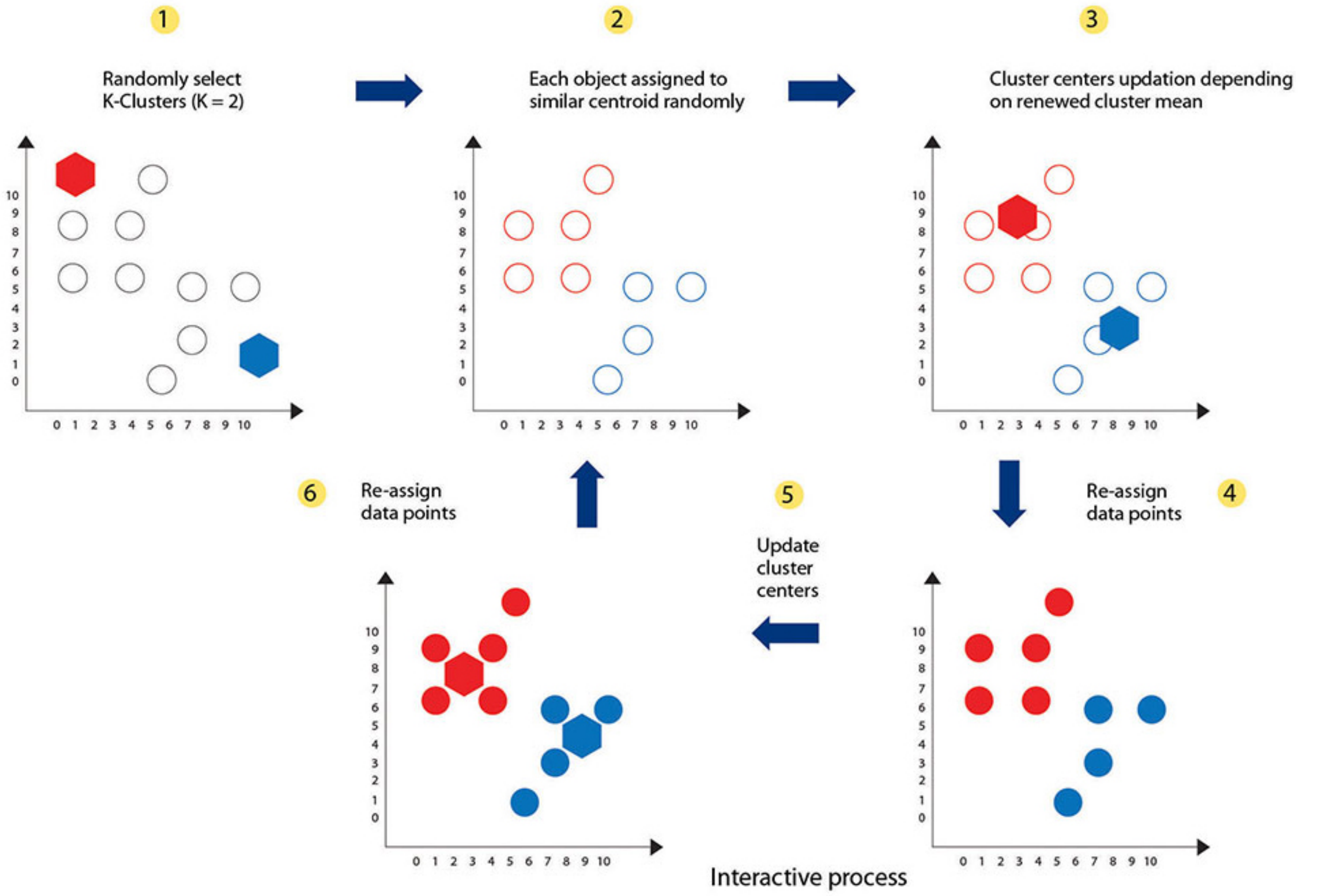
Clustering Algorithms & *k*

- → *k*-means: centroid-based
- → bi-secting: top-down
- → single-linkage (SL), complete-linkage (CL), average-linkage (AL): bottom-up



Source: https://homepages.uc.edu/~niunn/papers/RE12.pdf

k-means (e.g., *k*=2)



1 Randomly select K-Clusters (K = 2)

2 Each object assigned to similar centroid randomly

3 Cluster centers updation depending on renewed cluster mean

4 Re-assign data points

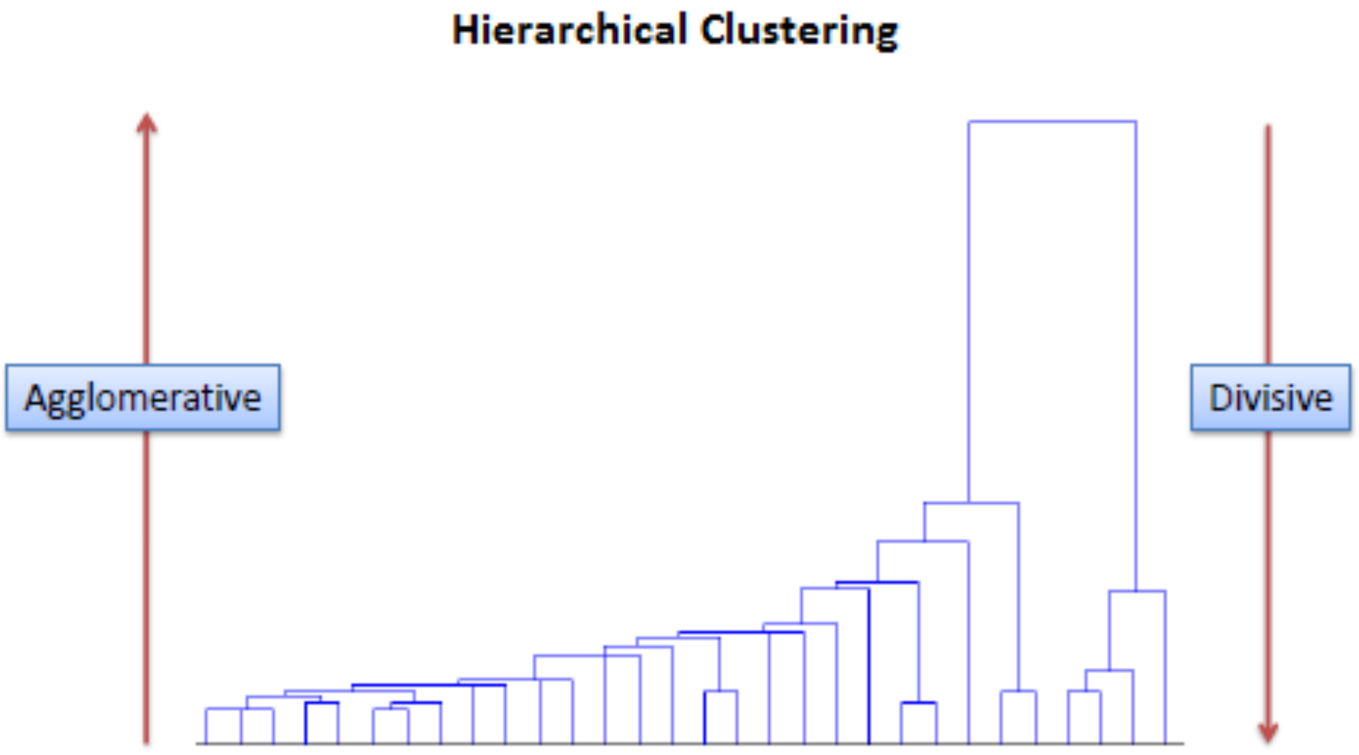
5 Update cluster centers

6 Re-assign data points

Interactive process

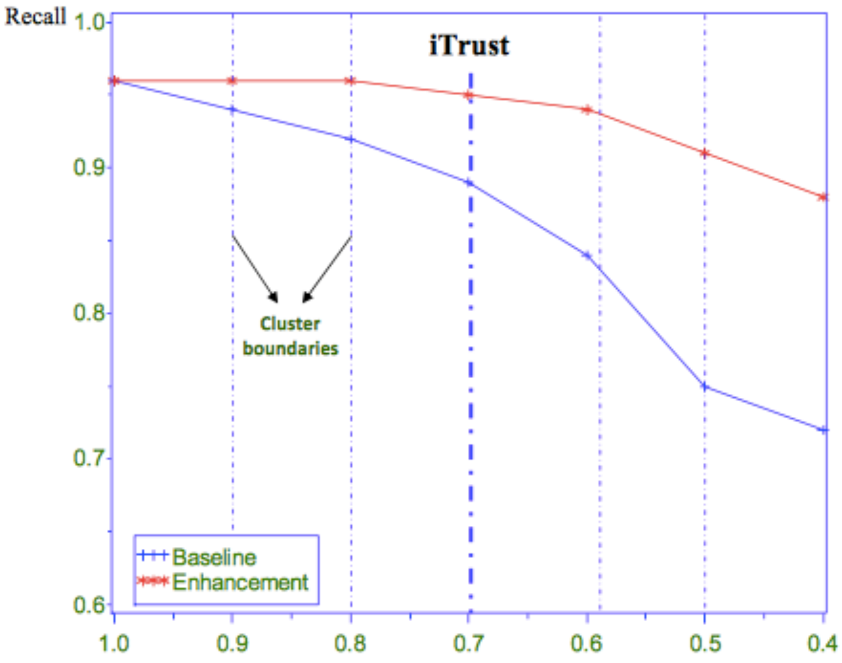
Source: https://www.brandidea.com/kmeans.html

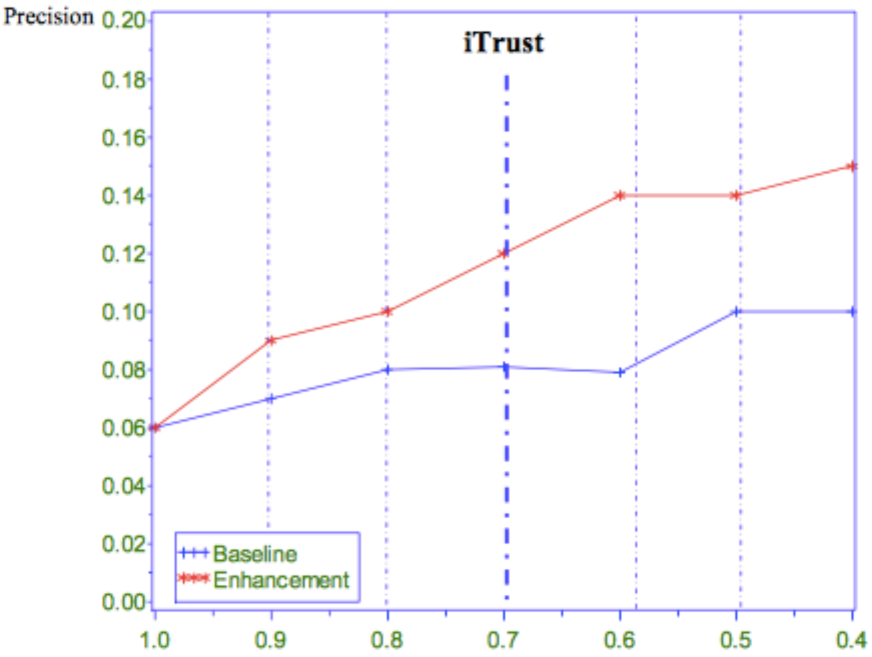
Hierarchical Clustering



Source: *https://www.saedsayad.com/clustering_hierarchical.htm*

How much better?

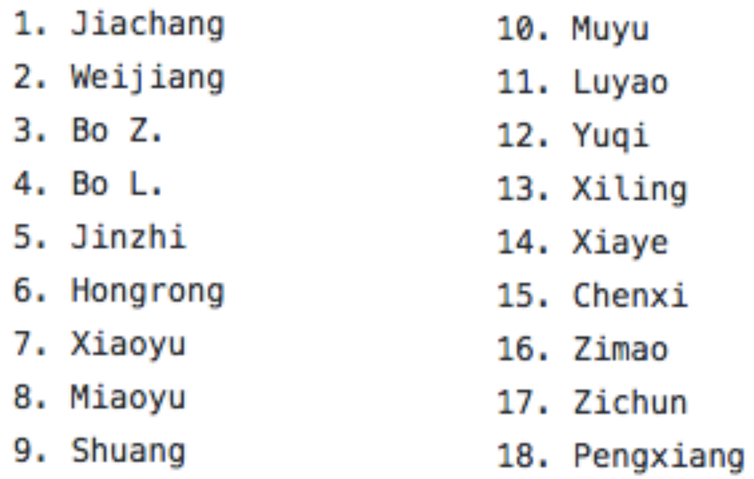




ASN4 grader program

RE Story (ASN2) Schedule

- Check 'ASN2-Schedule' on the course website
- We'll start at 9am on Monday (July 26)
- Each presenter has 5-10 minutes
- All the students are required to attend all the presentations



```
1. Jiachang
2. Weijiang
3. Bo Z.
4. Bo L.
5. Jinzhi
6. Hongrong
7. Xiaoyu
8. Miaoyu
9. Shuang
10. Muyu
11. Luyao
12. Yuqi
13. Xiling
14. Xiaye
15. Chenxi
16. Zimao
17. Zichun
18. Pengxiang
```

Source: *Google Images*