



# Requirements Engineering (Summer 2021)

Prof. Nan Niu ([nan.niu@uc.edu](mailto:nan.niu@uc.edu))

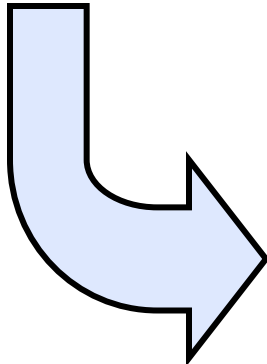
<https://github.com/nanniu/RE-Summer2021>



# Today's Menu

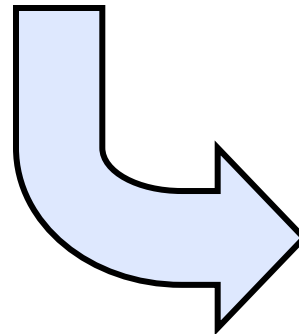
Wednesday (July 21)

RE Research  
(ASN2, ASN3 Q&A)



Thursday (July 22):

Req.s Traceability  
ASN4 Release



Friday (July 23):

Unsupervised Learning  
(ASN4 Q&A)



## Functional vs. Nonfunctional

- Functional requirements describe WHAT the software does
- Nonfunctional requirements (NFRs) describe HOW WELL the software does it
- Eliciting NFRs



# Elicitation Techniques

## → Traditional techniques

- ↳ Introspection
- ↳ Reading existing documents
- ↳ Analyzing hard data
- ↳ Interviews
  - Open-ended
  - Structured
- ↳ Surveys / Questionnaires
- ↳ Meetings

## → Collaborative techniques

- ↳ Group techniques
  - Focus Groups
  - Brainstorming
- ↳ JAD/RAD workshops
- ↳ Prototyping
- ↳ Participatory Design

## → Cognitive techniques

- ↳ Task Analysis
- ↳ Protocol Analysis
- ↳ Knowledge Acquisition Techniques
  - Card Sorting
  - Laddering
  - Repertory Grids
  - Proximity Scaling Techniques

## → Contextual approaches

- ↳ Ethnographic Techniques
  - Participant Observation
  - Ethnomethodology
- ↳ Discourse Analysis
  - Conversation Analysis
  - Speech Act Analysis
- ↳ Socio-technical Methods
  - Soft Systems Analysis



focus

stakeholders in RE.....

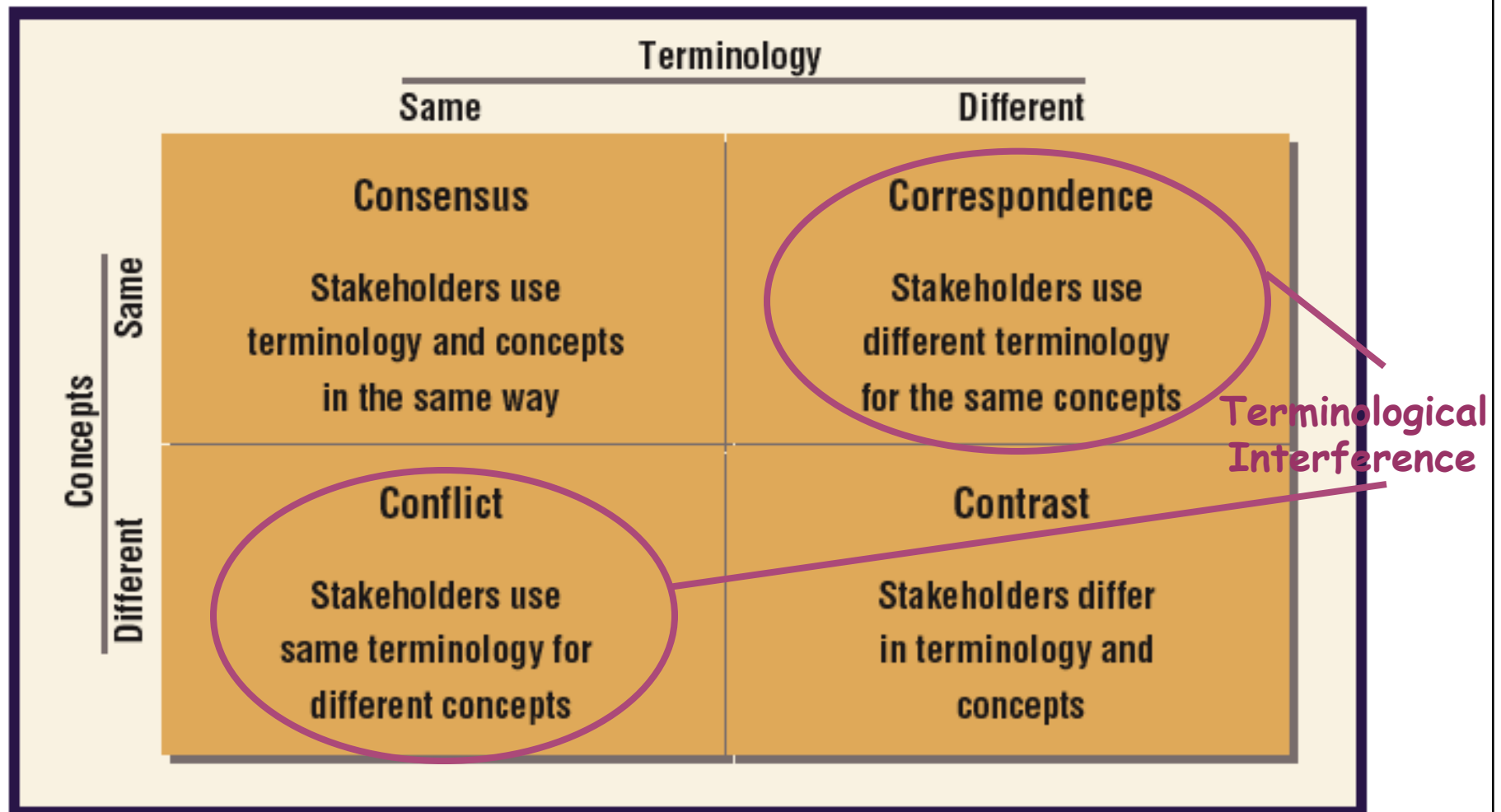
# **So, You Think You Know Others' Goals?**

## **A Repertory Grid Study**

**Nan Niu and Steve Easterbrook**, *University of Toronto*



# Concepts and Terminology





# Repertory Grid Technique (RGT)

- ⇒ George Kelly (1955), psychotherapy
- ⇒ verbalize how people construe certain factors within the area of interest
  - ↳ verbalizations: constructs (bipolar in nature)
  - ↳ factors: elements



# RGT Example

## ⇒ Information sources

↳ TV, Newspaper, Radio, NewsGroup, Web, etc.

↳ elements in RGT

## ⇒ Triad: (A) TV (B) Newspaper (C) NewsGroup

↳ construct: many focuses (A,B) vs. single focus (C)

↳ as a rating scale (1-5), and each element is assigned a rating on that construct





# Sample Repertory Grid

	TV	Newspaper	Radio	Newsgroup	...	
Many focuses	1	2	2	5	...	Single focus
Multimedia	1	4	2	5	...	Text
Entertaining	1	3	1	3	...	Not entertaining
Two-way	5	4	4	2	...	One-way
...	...	...	...	...	...	...



## Requirements Goal Models

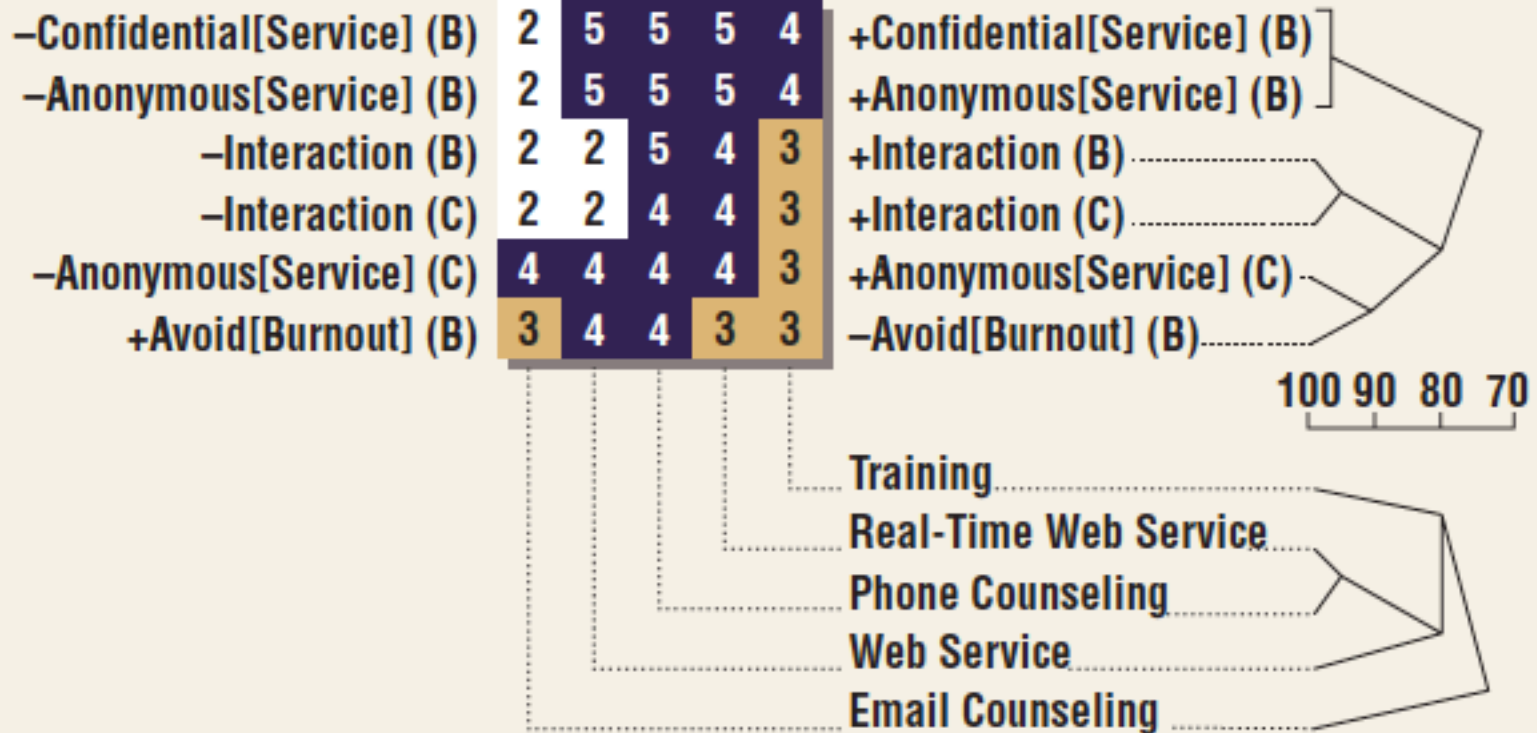
- ⇒ **Softgoals - Constructs - Unique to personal views**
- ⇒ **Tasks - Elements - Shared among stakeholders**
  
- ⇒ **Assume: people focusing on similar topics would agree on the definition of a common set of concrete tasks within the area of interest**
  
- ⇒ **Idea: compare stakeholder's constructs by how they relate to a shared set of concrete entities, rather than by any terms the stakeholders use to describe them**



# Kids Help Phone

Focus Grid Projection, Domain: KHP

Context: Counseling, 5 tasks, 6 softgoals



B - Bob C - Cem



## Observations

### ⇒ Trivial correspondence

↳ High-level softgoals about counseling: Good, Helpful, Proper, High-Quality, etc.

### ⇒ Numerical threshold

↳ Anonymous[Service] (Cem) versus (Bob)

### ⇒ Conflicts beyond terminological level

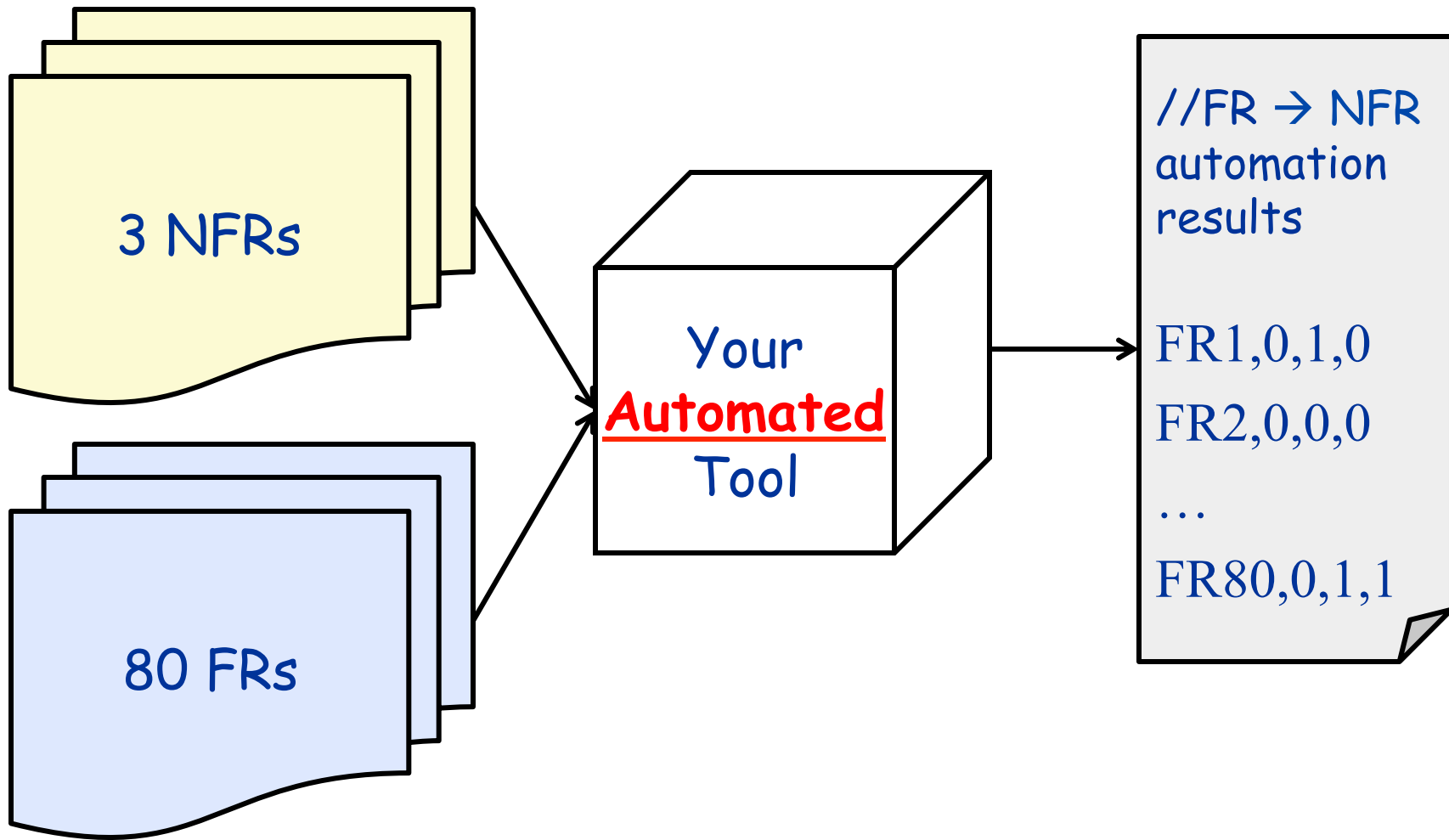
↳ (Ana) "Consult New Technique" would "Make-Difficult[Work]", hence hurt "Avoid[Burnout]"

↳ (Bob) "Consult New Technique" could help "High[Morale]", thus help "Avoid[Burnout]"

### ⇒ This leads us to Assignment 4

↳ Linking FRs and NFRs automatically

## ASN4: A conceptual picture

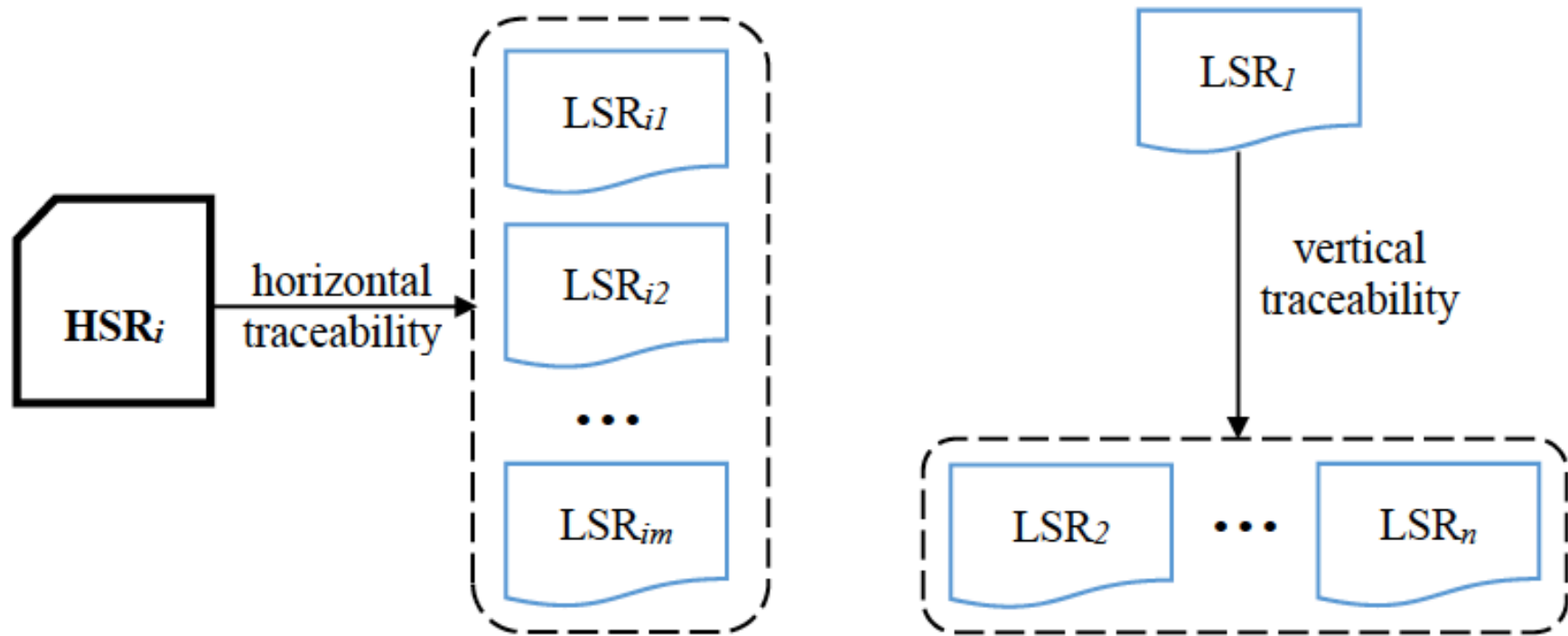


# What's "req.s traceability"?

⇒ the ability to describe and follow the life of a requirement

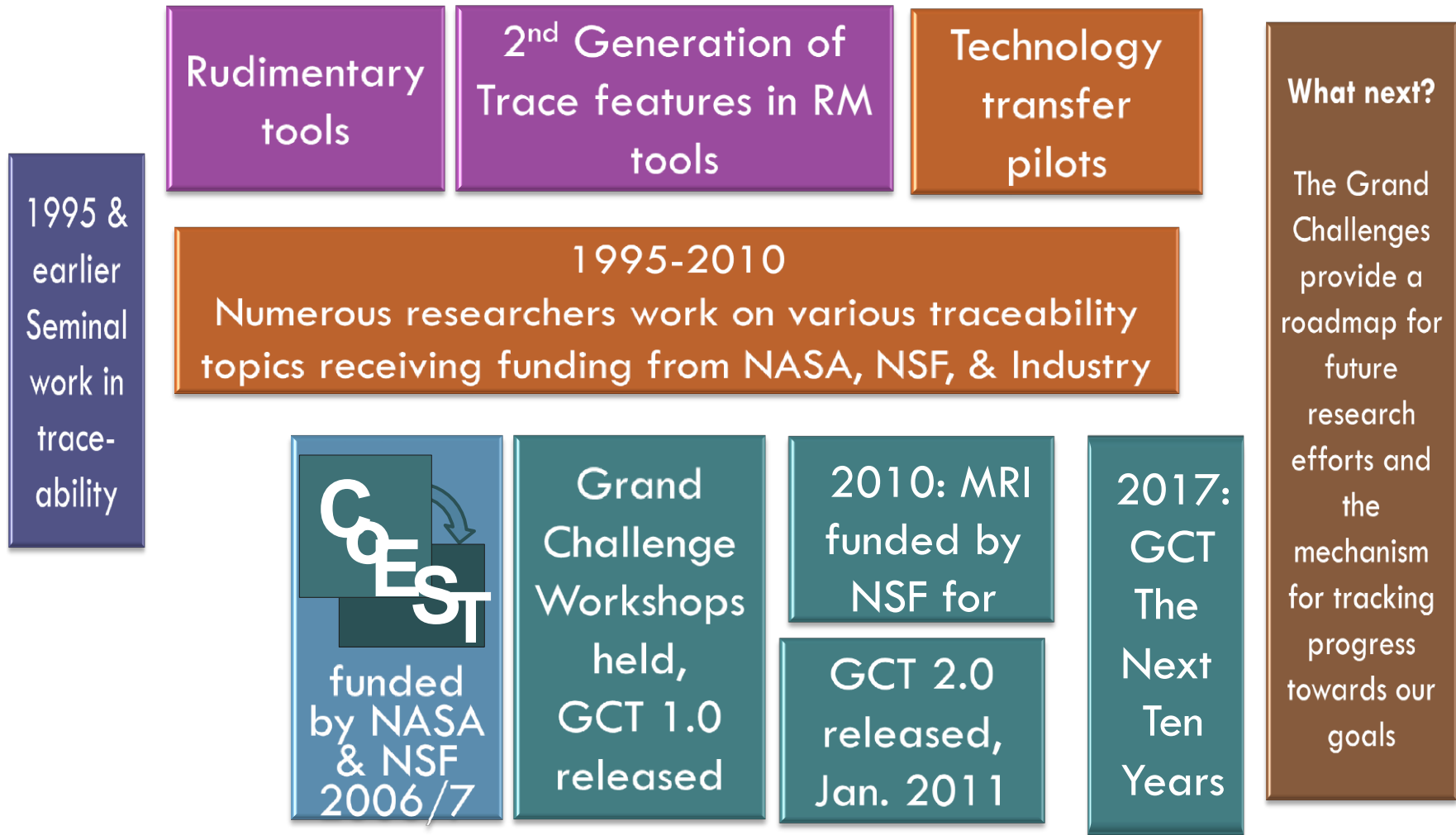


# Horizontal and vertical traceability





# Roadmap of "req.s traceability"



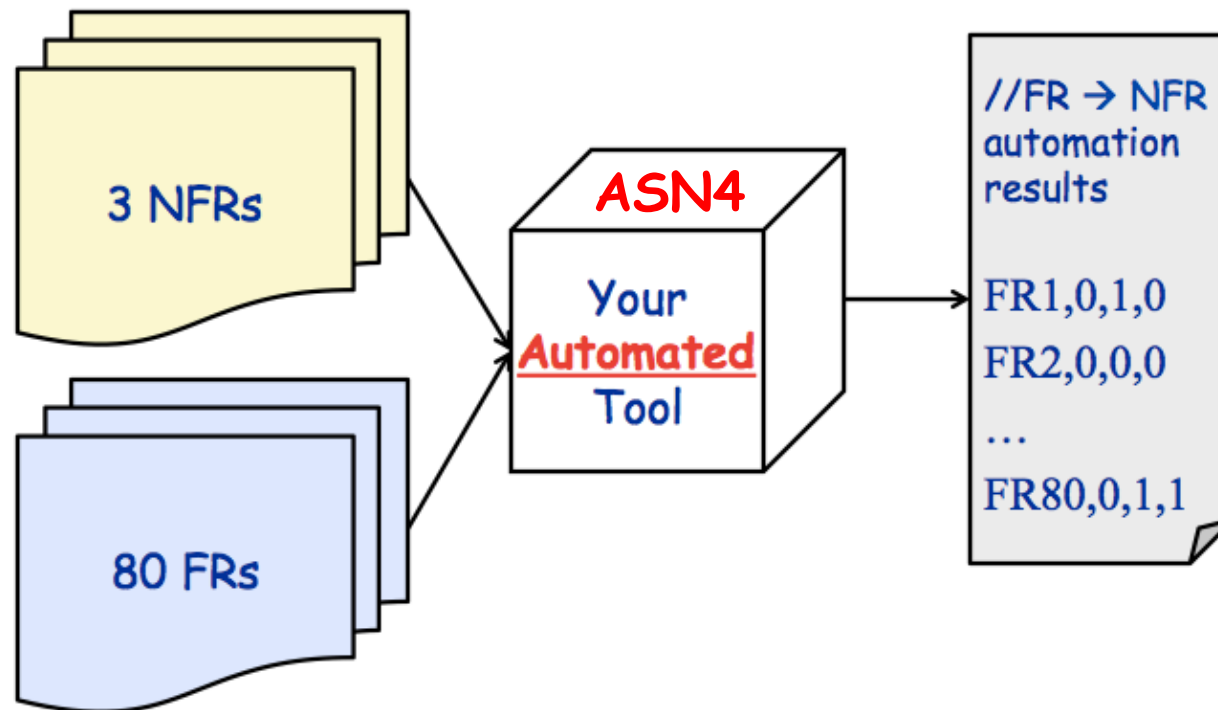


## Relating FRs to NFRs

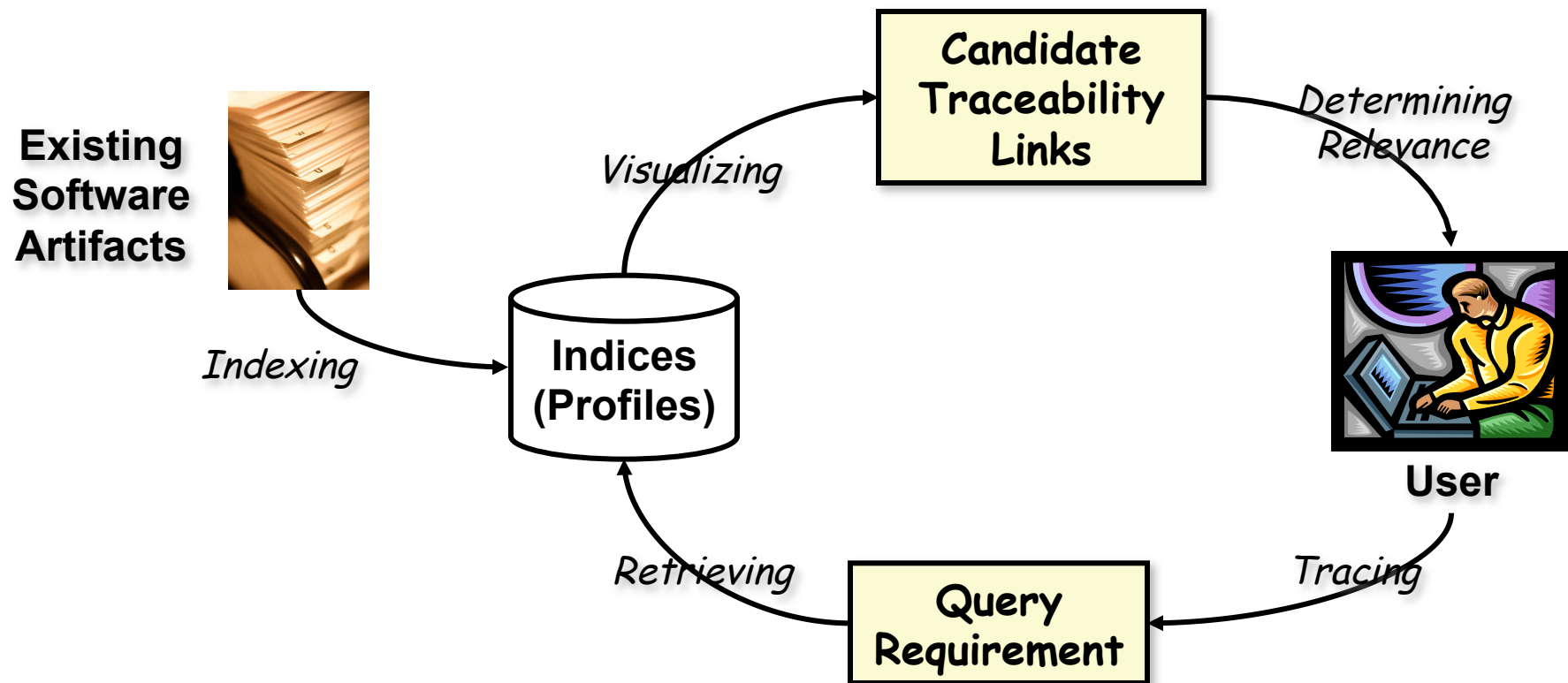
- ↪ **NFR1 (Operational):** The system shall interface with the Choice Parts System. This provides the feed of recycled parts data.
- ↪ **NFR2 (Usability):** Users shall feel satisfied using the system. 85% of all users will be satisfied with the system.
- ↪ **NFR3 (Security):** Only adjusters can request recycled parts audit reports.
  
- ↪ **FR\_i:** The user shall search for the preferred repair facility using vehicle location and radius in miles.
- ↪ **FR\_j:** The estimator shall search for available recycled parts using damaged vehicle parts information.

# Correct Answers

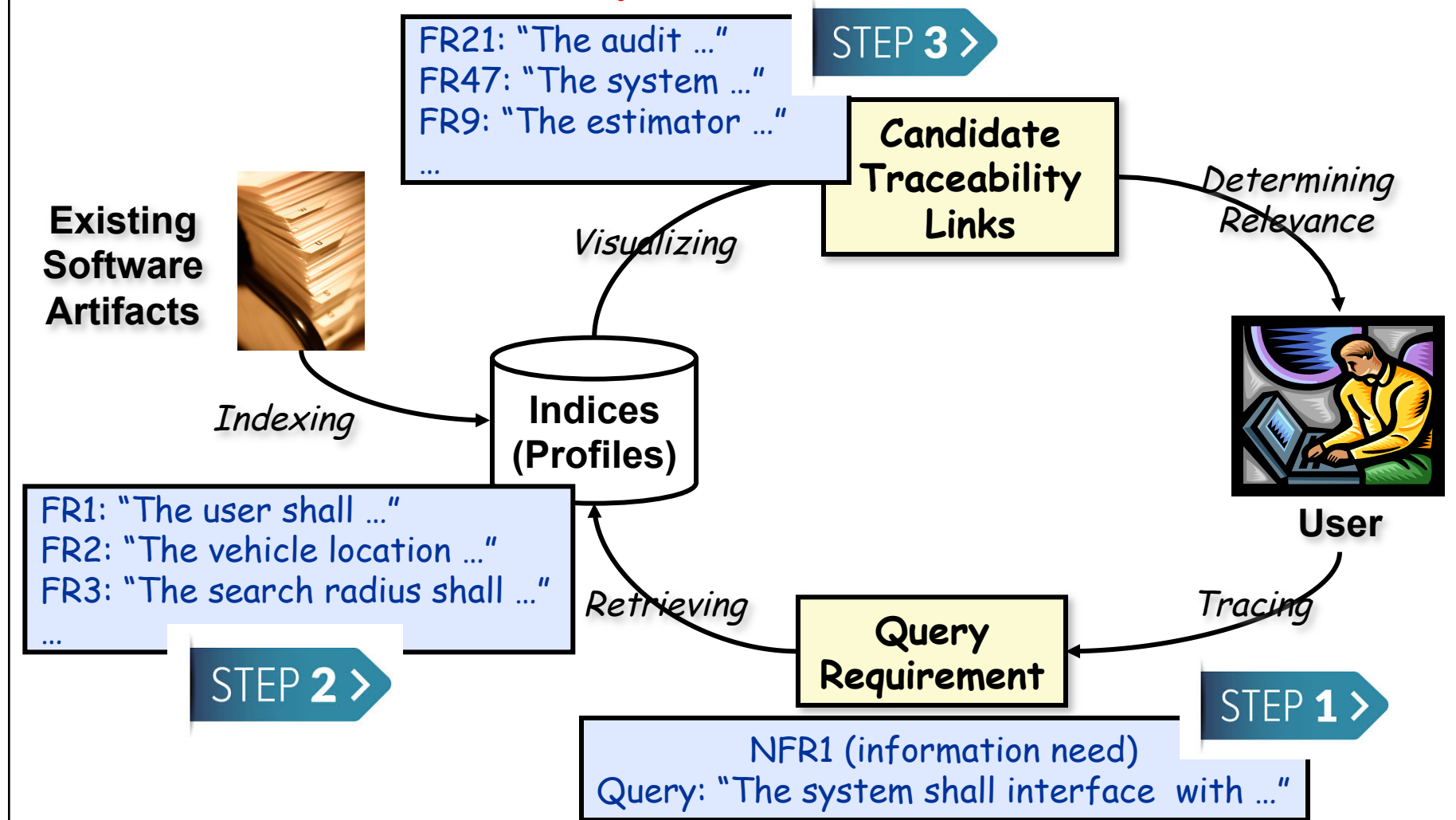
	NFR1 (Operational)	NFR2 (Usability)	NFR3 (Security)
FR <sub>i</sub>	0	1	0
FR <sub>j</sub>	1	1	0



# IR-Based Requirements Traceability



# IR-Based ASN4 Solution (fully automatic)



## Example

→ Two requirements

↳ r1 = "create and deactivate patients profile"

↳ r2 = "patients create and edit profile"

→ In this lecture, we introduce some basic retrieval methods: set-based, Jaccard, tf-idf.

→ Assumption of IR-based **ASN4** solution

↳ the more textual similarity there is between the two requirements, the more likely one is linked with (traceable to) the other

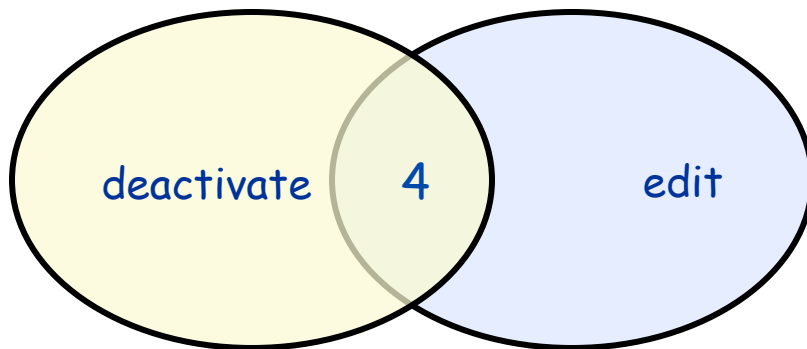
# Similarity based on set overlapping

→ Basic formula

$$S(R1, R2) = \frac{2 |R1 \cap R2|}{(|R1| + |R2|)}$$

↪ r1 = "create and deactivate patients profile"

↪ r2 = "patients create and edit profile"



→ Resulting similarity

↪  $S(r1, r2) = (2 \times 4) / (5 + 5) = 0.8$

↪ Suppose the threshold is 0.5, then {r1, r2} would be regarded as traceable to each other



# Similarity based on Jaccard index

→ Basic formula      The Jaccard similarity coefficient,  $J$ , is given as

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}.$$

$M_{11}$  represents the total number of attributes where  $A$  and  $B$  both have a value of 1.

$M_{01}$  represents the total number of attributes where the attribute of  $A$  is 0 and the attribute of  $B$  is 1.

$M_{10}$  represents the total number of attributes where the attribute of  $A$  is 1 and the attribute of  $B$  is 0.

$M_{00}$  represents the total number of attributes where  $A$  and  $B$  both have a value of 0.

→ In our example

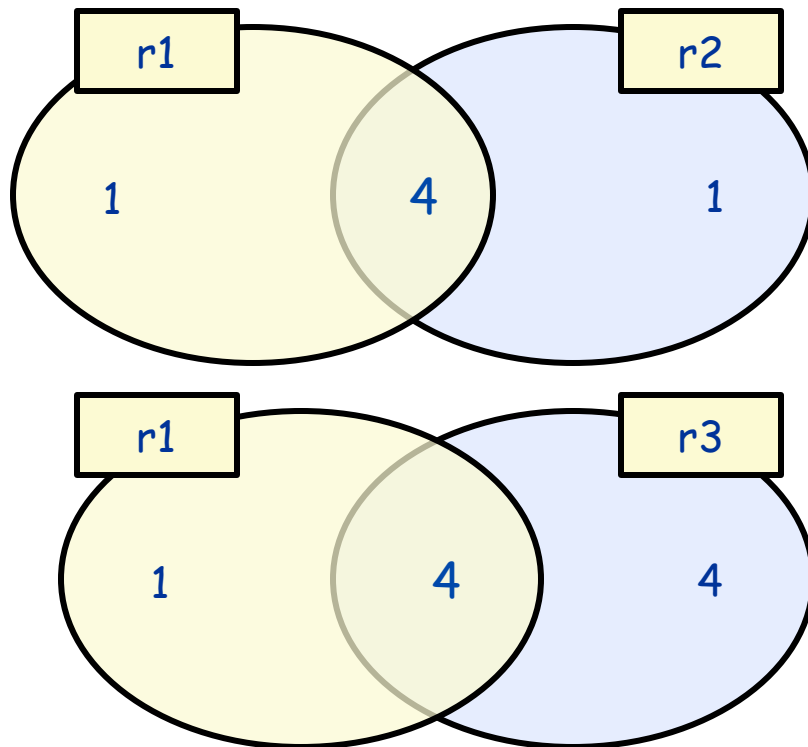
	create	and	deactivate	patients	profile	edit	a	including	photo
r1	1	1	1	1	1	0	0	0	0
r2	1	1	0	1	1	1	0	0	0
r3	1	1	0	1	1	1	1	1	1

## Jaccard (cont'd)

↪  $r1$  = "create and deactivate patients profile"

↪  $r2$  = "patients create and edit profile"

↪  $r3$  = "patients create and edit profile including a photo"



→ Set-based similarity

$$\hookrightarrow S(r1, r2) = (2 \times 4) / (5 + 5) = 0.8$$

$$\hookrightarrow S(r1, r3) = (2 \times 4) / (5 + 8) = 0.62$$

→ Jaccard-based similarity

$$\hookrightarrow S(r1, r2) = 4 / 6 = 0.67$$

$$\hookrightarrow S(r1, r3) = 4 / 9 = 0.44$$





## Results So Far (threshold=0.5)

→ Our example

↪ r1 = "create and deactivate patients profile"

↪ r2 = "patients create and edit profile"

↪ r3 = "patients create and edit profile including a photo"

↪ r4 = "patients create and and edit edit edit profile"

→ Set-based overlap

2,4	(1.00)
1,2	(0.80)
1,4	(0.80)
2,3	(0.77)
3,4	(0.77)
1,3	(0.62)

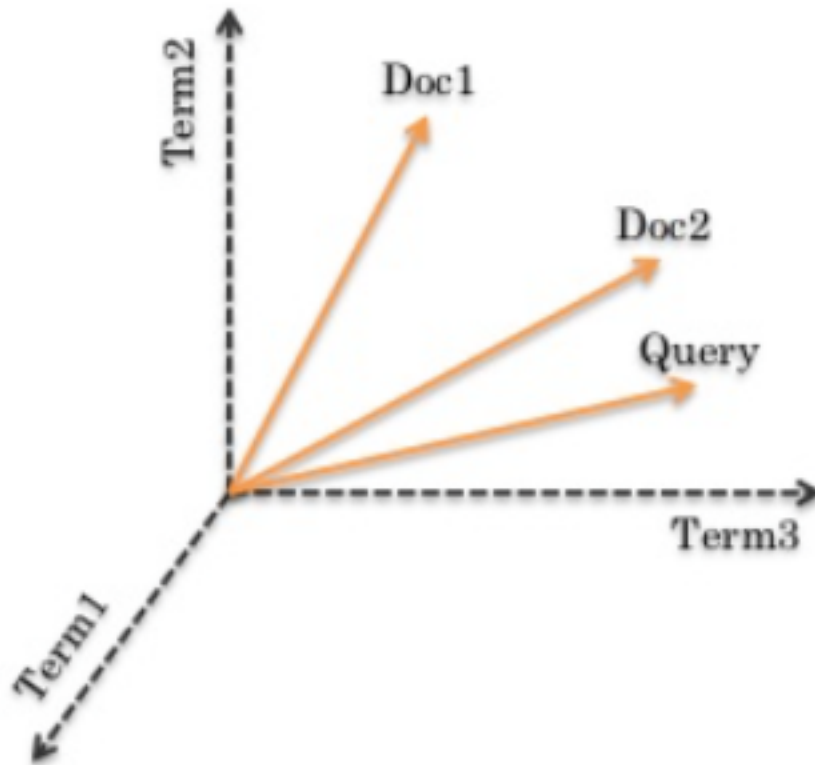
→ Jaccard index

2,4	(1.00)
1,2	(0.67)
1,4	(0.67)
2,3	(0.63)
3,4	(0.63)

→ tf-idf

2,4	(0.89)
1,2	(0.56)
2,3	(0.50)
3,4	(0.45)
1,4	(0.36)

# VSM (vector space model)



$$\text{Cos}(D_i, Q) = \frac{\sum_{j=1}^t d_{ij} * q_j}{\sqrt{\sum_{j=1}^t d_{ij}^2 * \sum_{j=1}^t q_j^2}}$$



## tf-idf

	create	and	deactivate	patients	profile	edit	a	...	photo
r1	1	1	1	1	1	0	0	...	0
r2	1	1	0	1	1	1	0	...	0
r3	1	1	0	1	1	1	1	...	1
r4	1	2	0	1	0	3	0	...	0

$$sim(d, q) = \cos(d, q) = \frac{\sum_{i=1}^N w_i \cdot q_i}{\sqrt{\sum_{i=1}^N w_i^2 \cdot \sum_{i=1}^N q_i^2}}$$

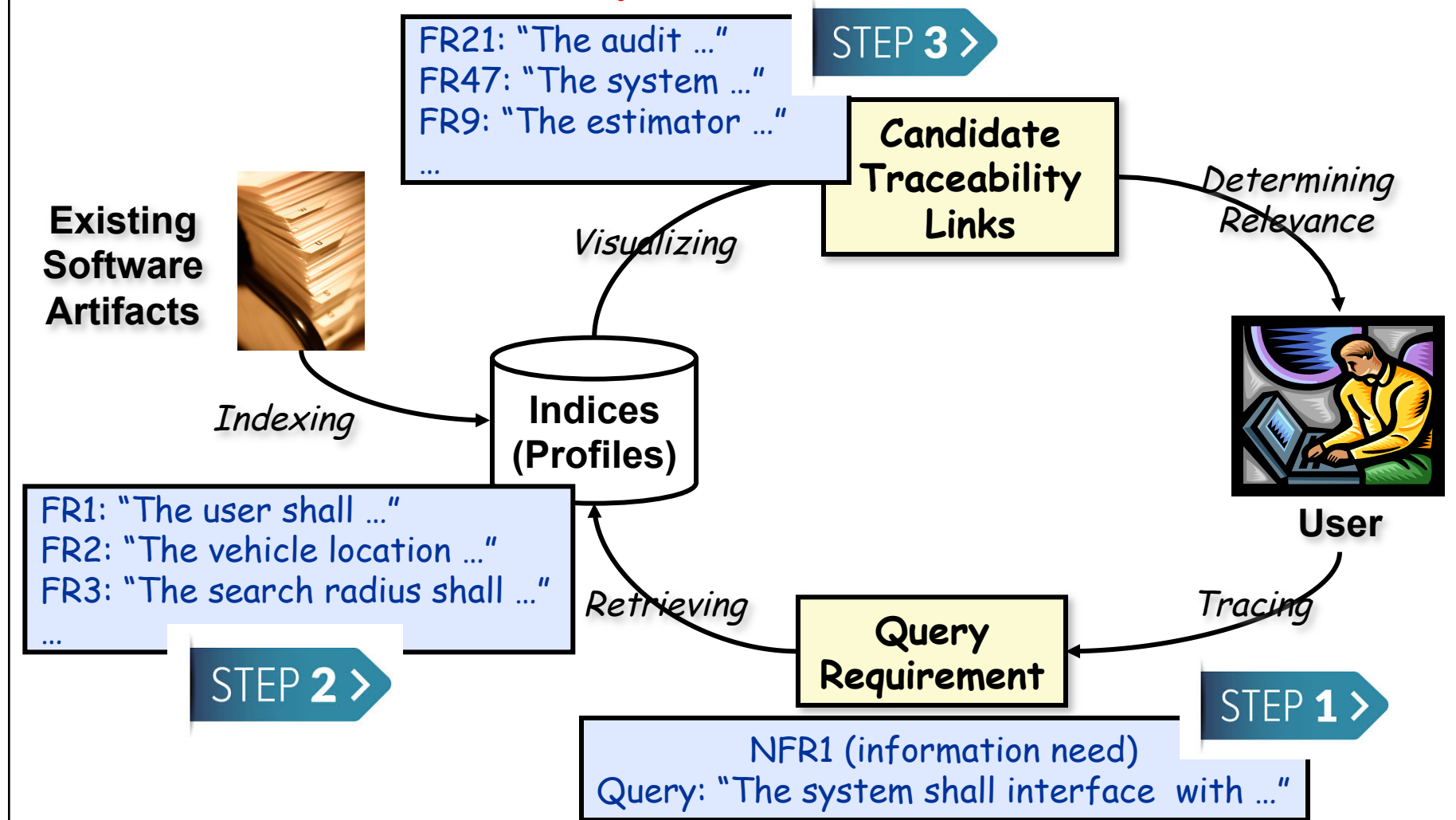
$$w_i = tf_i(d) \cdot idf_i,$$

$$idf_i = \log_2 \left( \frac{n}{df_i} \right),$$

*sim(r2, r4) =*

$[1 * \log(4/4+1)] * [1 * \log(4/4+1)]$  //create  
 $+ [1 * \log(4/4+1)] * [2 * \log(4/4+1)]$  //and  
 $+ [1 * \log(4/4+1)] * [1 * \log(4/4+1)]$  //patients  
 $+ [1 * \log(4/3+1)] * [0 * \log(4/3+1)]$  //profile  
 $+ [1 * \log(4/3+1)] * [3 * \log(4/3+1)]$  //edit  
 ... // denominator  
 = 0.89

# IR-Based ASN4 Solution (fully automatic)





## When NFR1 is concerned, different ASN4 solutions will output different FRs

### Ana's tool outputs

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

### Bob's tool outputs

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

### Chris's tool outputs

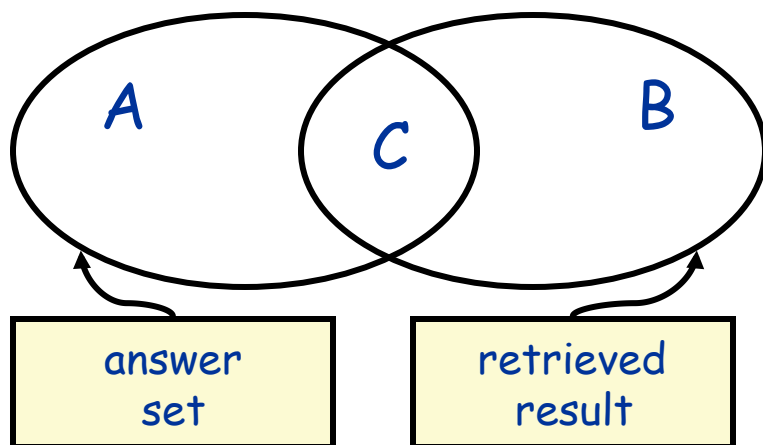
FR9: "The estimator ..."  
FR5: "The adjuster ..."  
FR11: "The system ..."  
FR47: "The system ..."  
FR76: "The system ..."

### David's tool outputs

FR9: "The estimator ..."  
FR5: "The adjuster ..."  
FR11: "The system ..."  
FR47: "The system ..."  
FR76: "The system ..."  
FR52: "The system ..."  
FR65: "The system ..."  
FR80: "The estimator ..."

**Ana's outputs = Bob's outputs**  
**Chris's \subset\_of David's**

## IR Metrics



Precision (accuracy) =  $|C| / |B|$

Recall (coverage) =  $|C| / |A|$

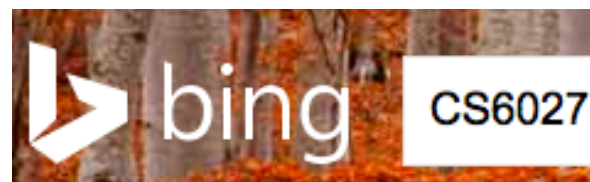
F-measure =  $\frac{(1+\beta^2) \times (P \times R)}{(\beta^2 \times P + R)}$

(F2-measure:  $\beta=2$ ; weights R twice as much as P)

Google

CS6027

Suppose  $|A|=10$



Result	Relevant
gr1	Yes
gr2	No
gr3	Yes
gr4	Yes
gr5	No

Precision<sub>Google</sub> =  $3 / 5 = 60\%$

Recall<sub>Google</sub> =  $3 / 10 = 30\%$

F2<sub>Google</sub> = 0.33

Precision<sub>Bing</sub> =  $2 / 3 = 67\%$

Recall<sub>Bing</sub> =  $2 / 10 = 20\%$

F2<sub>Bing</sub> = 0.23

Result	Relevant
br1	Yes
br2	Yes
br3	No



## Let's help David to decide the threshold

Rank	Candidate FR	Relevant?	Recall	Precision	F2
1	FR9	1	$1/20=0.05$	$1/1=1.00$	0.062
2	FR5	0	$1/20=0.05$	$1/2=0.50$	0.061
3	FR11	1	$2/20=0.10$	$2/3=0.67$	0.121
4	FR47	1	$3/20=0.15$	$3/4=0.75$	0.179
5	FR76	0	$3/20=0.15$	$3/5=0.60$	0.176
6	FR52	1	$4/20=0.20$	$4/6=0.67$	0.233
7	FR65	1	$5/20=0.25$	$5/7=0.71$	0.287
8	FR80	0	$5/20=0.25$	$5/8=0.63$	0.284

# ASN4 Given (Existing) Data

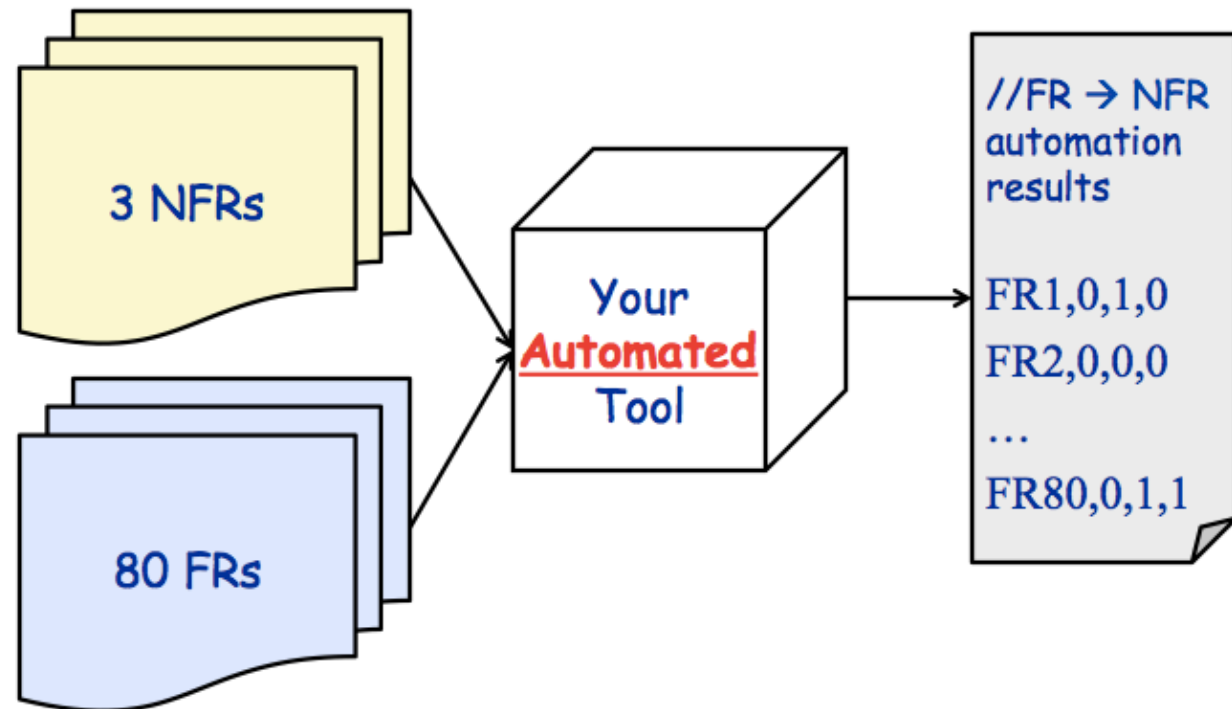
→ Answer set: trace-3nfr-80fr.txt

↪ FR1,0,1,0

↪ ...

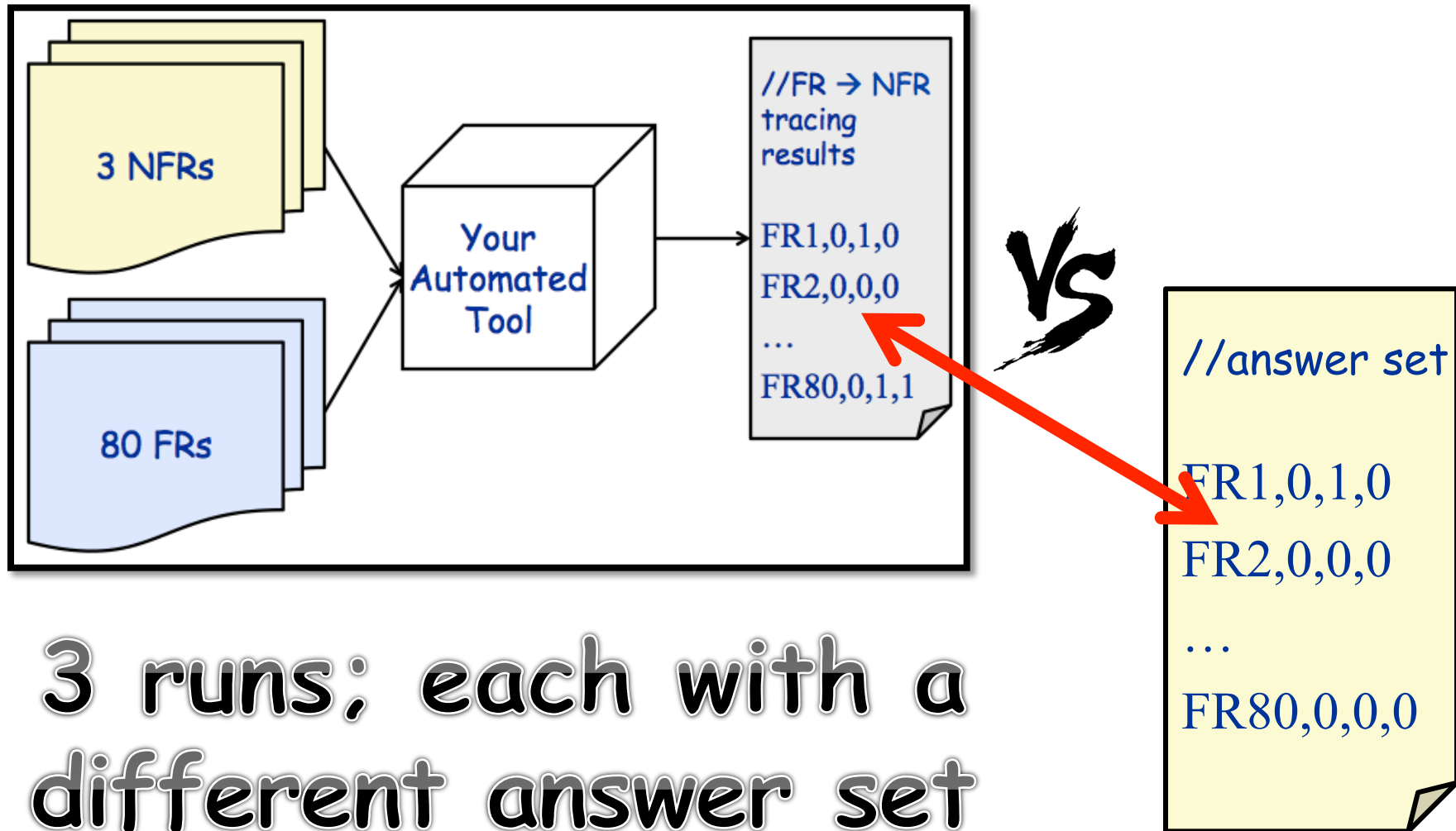
↪ FR9,1,1,0

↪ ...





# ASN4 Performance Evaluation





## Evaluating Your ASN4 Solution

→ The output of your ASN4 algorithm will be assessed via IR metrics

↳ Recall, Precision, and F2

→ Your ASN4 algorithm will be run three times in your 10-minute slot (July 28)

↳ 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)

# ASN4 Grading: **Absolute** Criteria

(Recall  $\geq 70\%$ ) and (Precision  $\geq 30\%$ )

- You'll receive 10 points if all of your 3 runs satisfy the above performance criteria
- You'll receive 7 points if 2 of your runs are good
- You'll receive 4 points if 1 of your runs is good
- You'll receive 1 point if you do your demo-grading

IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 32, NO. 1, JANUARY 2006

## Advancing Candidate Link Generation for Requirements Tracing: The Study of Methods

Jane Huffman Hayes, *Member, IEEE Computer Society*, Alex Dekhtyar, and  
Senthil Karthikeyan Sundaram, *Student Member, IEEE*



# Paper in the 'Readings' Content Area

TABLE 3  
Classification of Results and Relationship between Measures  
and Requirements

Measure	Acceptable	Good	Excellent
Recall	60% — 69%	70% — 79%	80% — 100%
Precision	20% — 29%	30% — 49%	50% — 100%
Lag	3 — 4	2 — 3	0 — 2





## ASN4 Grading: **Relative** Criteria

↳ 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)

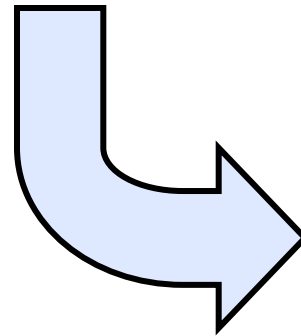
$$| F2(\text{run2}) - F2(\text{run1}) | / F2(\text{run1}) < 10\%$$

$$| F2(\text{run3}) - F2(\text{run2}) | / F2(\text{run2}) < 10\%$$

- You'll receive 10 points if both relative criteria hold
- You'll receive 6 points if only one criterion holds
- You'll receive 2 points if you do your demo-grading



# Tomorrow, we're heading to ...



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

