MANCHESTER 1824
School of Computer Science
The University of Manchester

Using Frame Embeddings to Identify Semantically Related Software Requirements

Waad Alhoshan, Riza Batista-Navarro and Liping Zhao

Semantic Relatedness in RE

- Detecting related NL requirements is **tricky** sometimes!
  - Inherent problems due to NL, e.g. ambiguity and incompleteness
  - Writing in NL does not adhere to any formalism

- Req-1: The transaction records are kept into a central database of the Bank and only authorised users are able to view the documents.
- Req-2: The Bank's reports are stored and restricted i.e. accessing the logs should be allowed to specific users.
- Req-3: The Bank's clients are requested to confirm their personal information regularly.
- Req-4: Every year the bank control system shall ask the clients to verify their contact information.

Semantic Frames

- **Semantic Frame** is defined as a coherent structure of concepts.
- **FrameNet** is an implementation of that theory:
  - More than 1200 frames.
  - Curated by language experts.
  - Frame contents: Definition, Core and non-core frame elements, lexical units and semantic relations with other frames (if any).

Semantic Frames Cont.

Req-1: The transaction records are kept into a central database of the Bank and only authorised users are able to view the documents.

FN-Req-1: The transaction **records** [*Records*] are **kept** [*Storing*] into a central database of the Bank and only **authorised** [*Deny_or_grant_permission*] users are **able** [*Capability*] to **view** [*Perception_active*] the **documents** [*Text*].

Req-2: The Bank's reports are stored and restricted i.e. accessing the logs should be allowed to specific users.

FN-Req-2: The Bank's **reports** [*Text*] are **stored** [*Storing*] and **restricted** [*Deny_or_grant_permission*] i.e. accessing the **logs** [*Records*] should be **allowed** [*Preventing_or_letting*] to **specific** [*Specific_individual*] users.

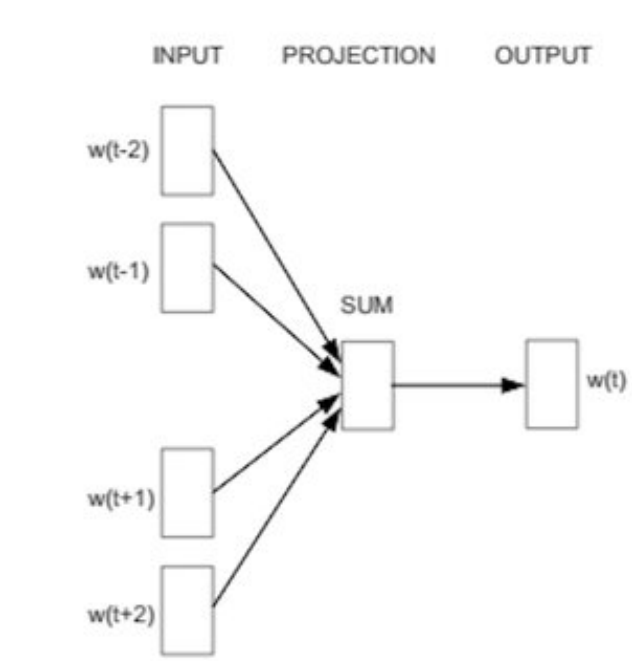
Req-3: The Bank's clients are requested to confirm their personal information regularly.

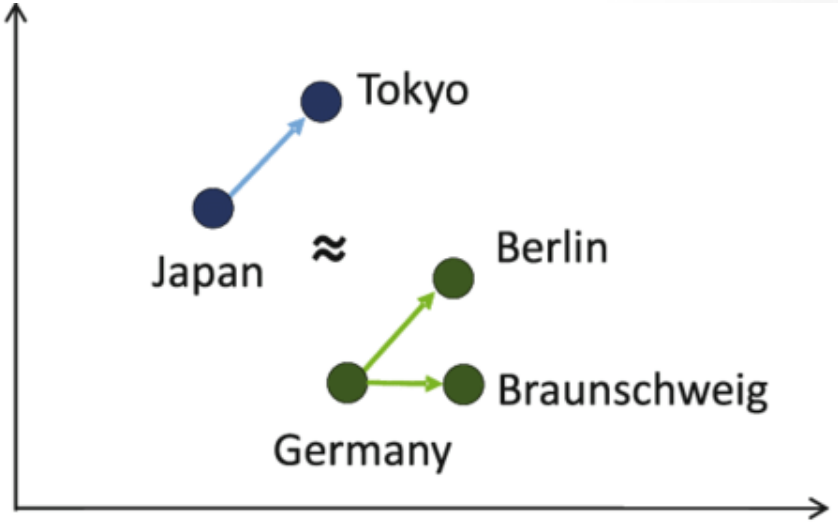
FN-Req-3: The Bank's clients are **requested** [*Request*] to **confirm** [*Verification*] their personal **information** [*Information*] **regularly** [*Frequency*].

Req-4: Every year the bank control systems shall ask the clients to verify their contact information.

FN-Req-4: **Every** [*Frequency*] **year** [*Calendric_unit*] the bank **control** [*Being_in_control*] **system** [System] shall **ask** [*Request*] the clients to **verify** [*Verification*] their **contact** [*Contacting*] **information** [*Information*].

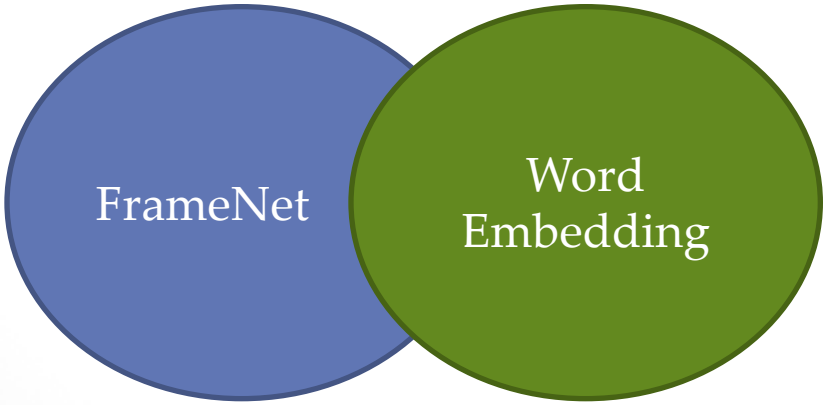
Word Embedding





NLP4RE

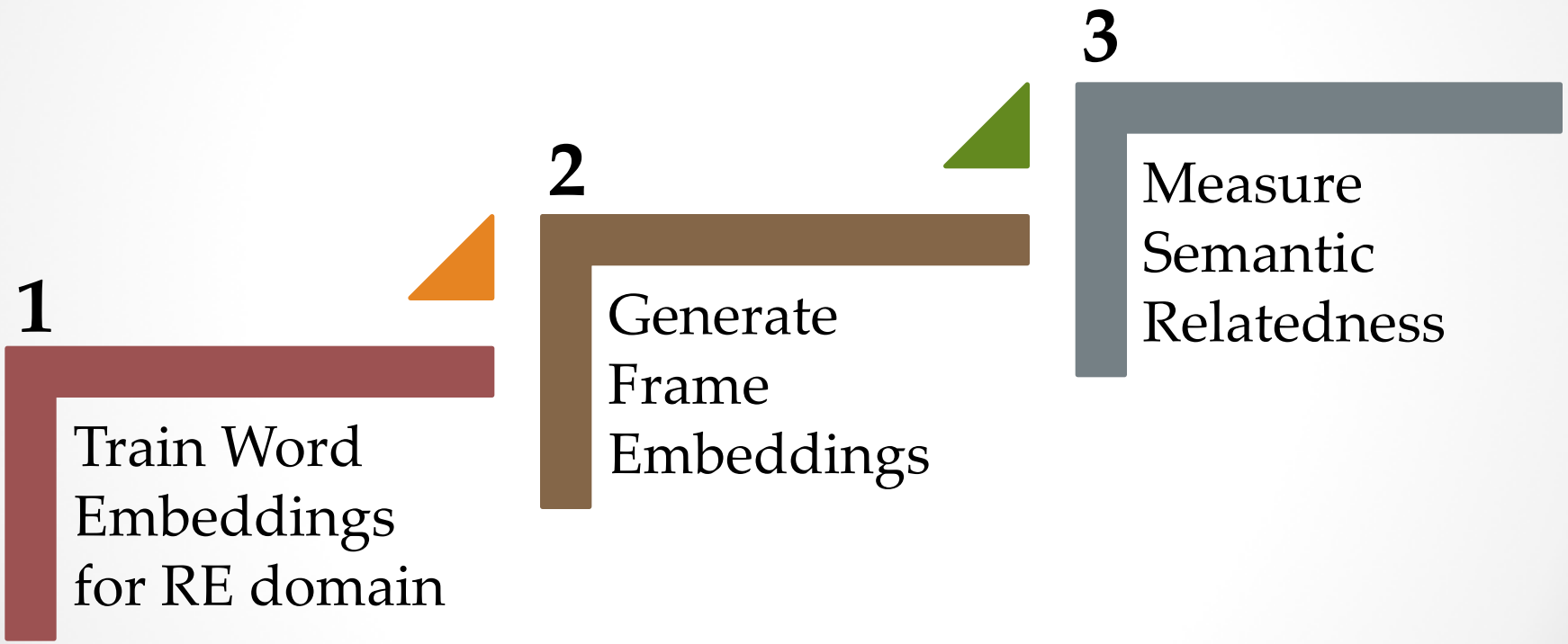
"Customizing general NLP techniques to make them applicable for solving the problems requirements engineers face in their daily practice." [1]



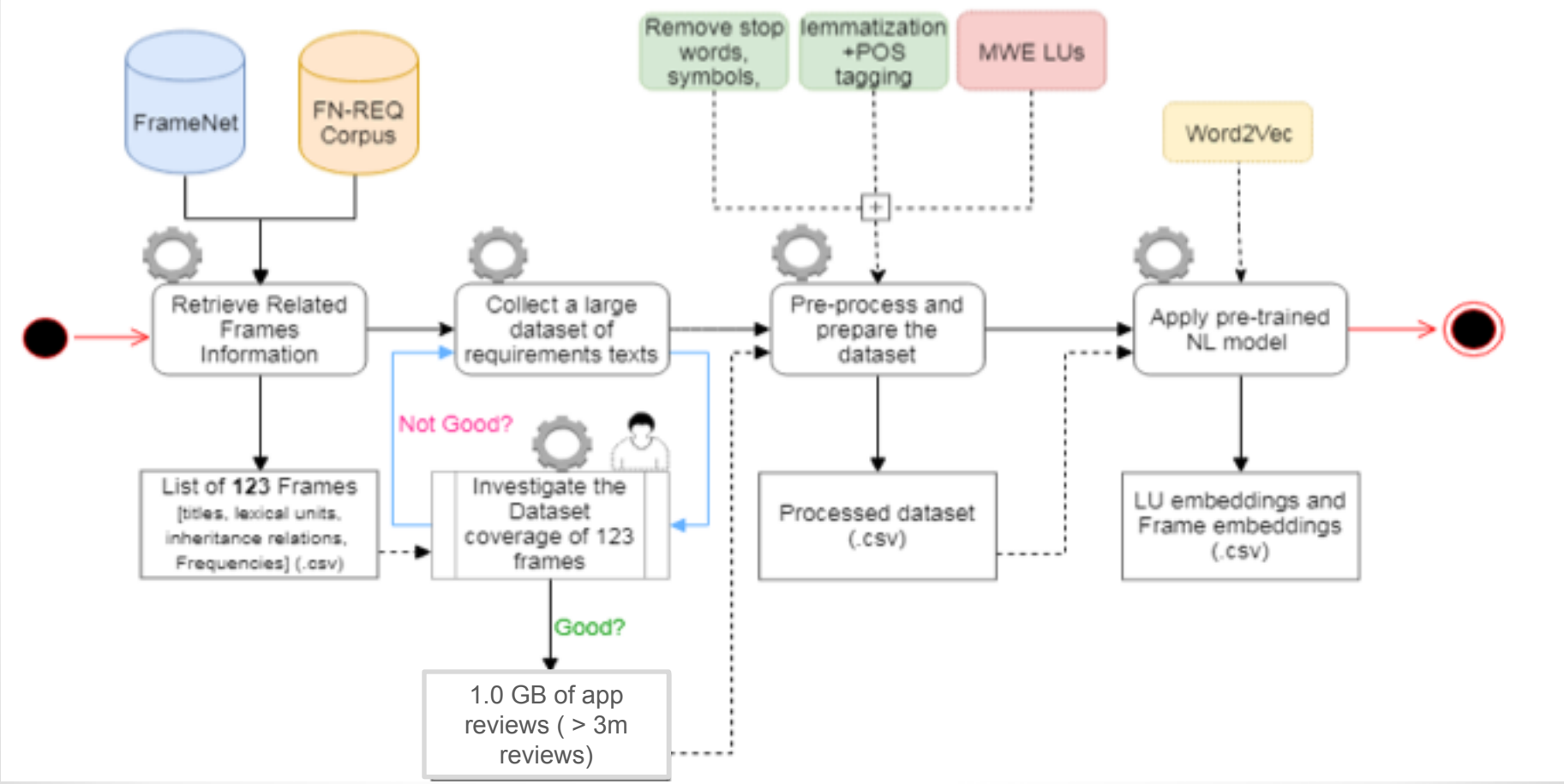
NLP4RE Cont.

- We published papers based on Corpus-based investigation research for using FrameNet in RE (FN-RE corpus) [2] [3].
- Also, to investigate ways for measuring Semantic Relatedness between Frames from RE perspective.
  - Knowledge-based measures (WUP and Path) [4]
  - Context-based measure(pre-trained word embedding for SE) [4] [5].

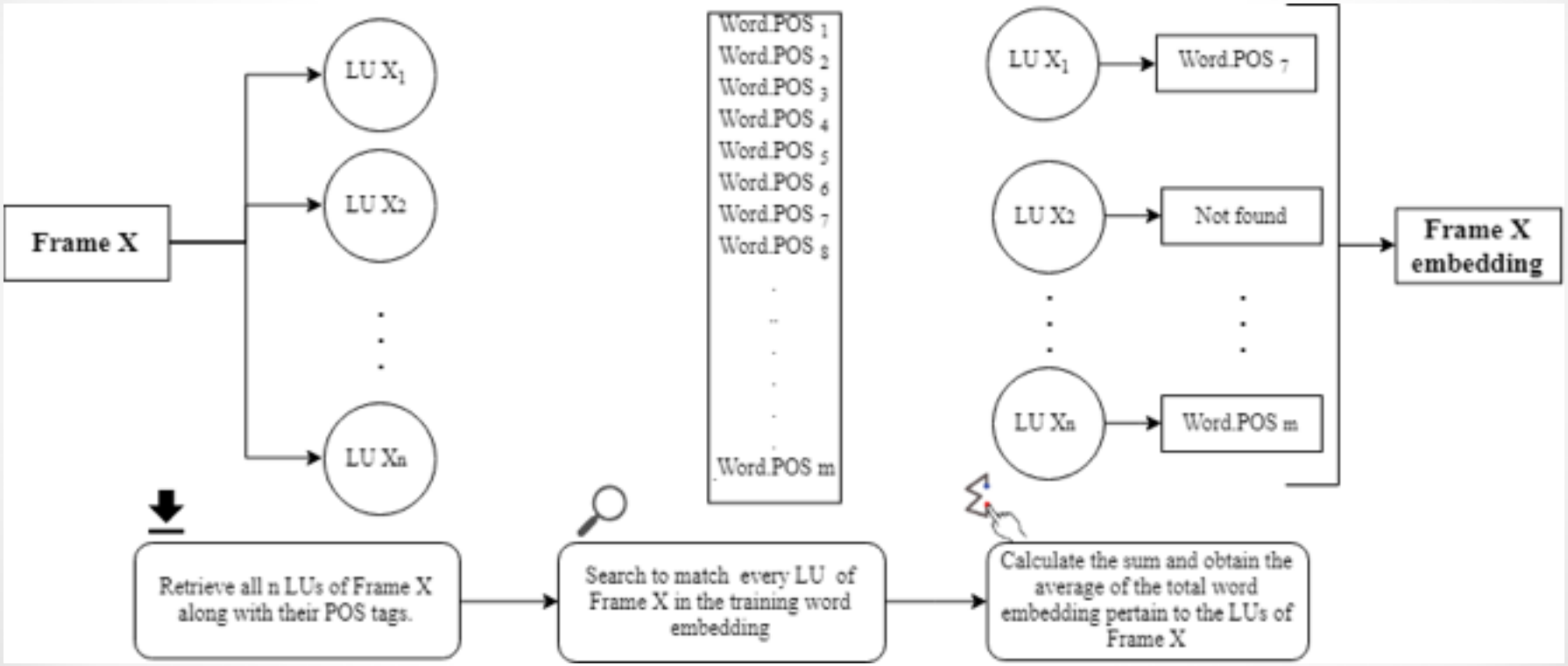
Frame-to-Frame Method



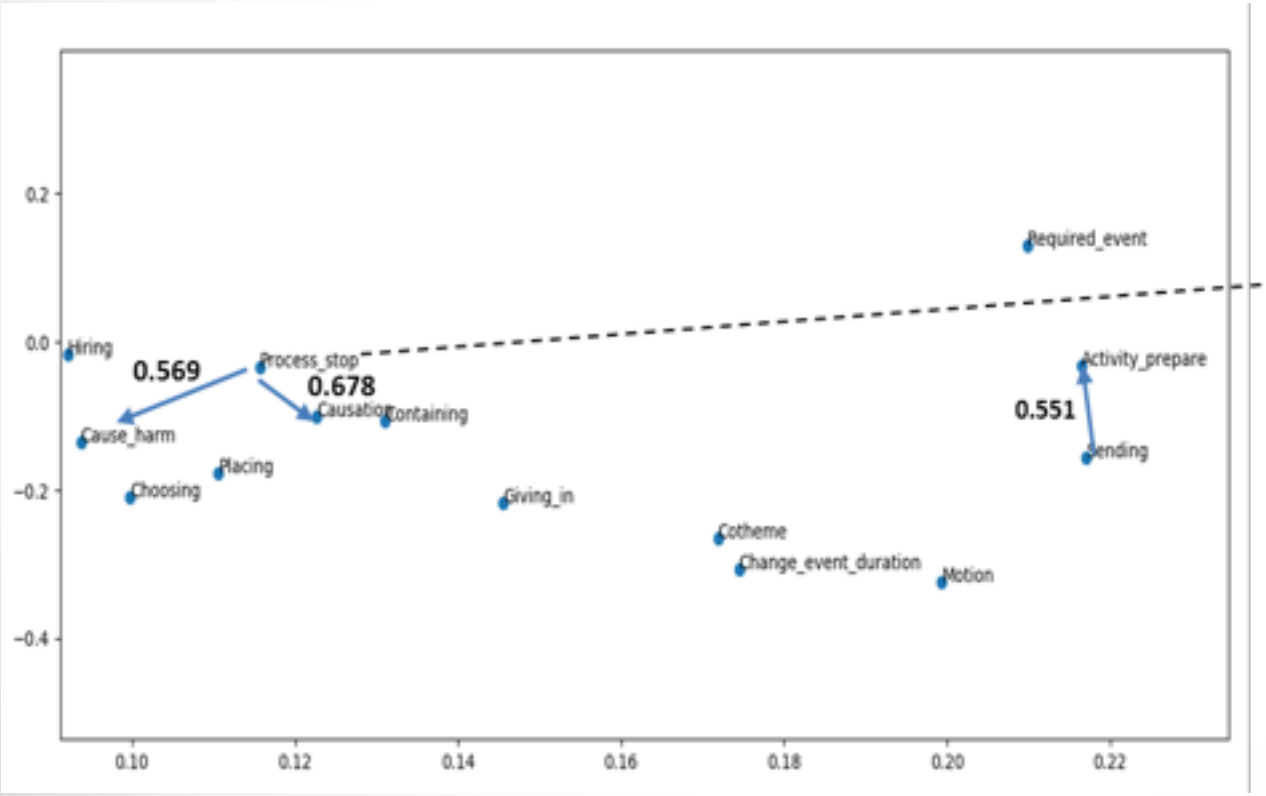
Step 1: Train Word Embedding



Step 2: Generate Frame Embedding



Step 2: Generate Frame Embedding Cont.



Top 10 similar frames for the frame 'Process_stop':

1. ('Change_operational_state', 0.6792)
2. ('Causation', 0.6787)
3. ('Preventing_or_letting', 0.6458)
4. ('Activity_start', 0.5976)
5. ('Deny_or_grant_permission', 0.5836)
6. ('Avoiding', 0.5768)
7. ('Placing', 0.5742)
8. ('Process_start', 0.5703)
9. ('Cause_harm', 0.5693)
10. ('Desiring', 0.5500)

Step 3: Measure Semantic Relatedness

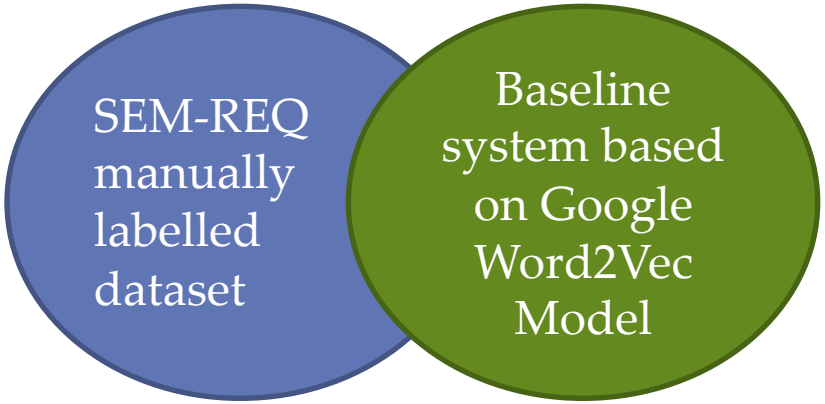
$$Ra = (Fa_1, Fa_2, \dots, Fa_n) \text{ and } Rb = (Fb_1, Fb_2, \dots, Fb_m) \qquad M = [Ra, Rb] = \begin{bmatrix} FR(Fa_1, Fb_1) & \cdots & FR(Fa_n, Fb_1) \\ \vdots & \cdots & \vdots \\ FR(Fa_1, Fb_m) & \cdots & FR(Fa_n, Fb_m) \end{bmatrix}$$

$$\vec{M} = \begin{bmatrix} (FR(Fa_1, Fb_1) & + \cdots + & FR(Fa_n, Fb_1))/n \\ \vdots & + \cdots + & \vdots \\ (FR(Fa_1, Fb_m) & + \cdots + & FR(Fa_n, Fb_m))/n \end{bmatrix} = \begin{bmatrix} \overrightarrow{FR}_1 \\ \vdots \\ \overrightarrow{FR}_n \end{bmatrix}$$

$$\downarrow M = \begin{bmatrix} (FR(Fa_1, Fb_1) & \cdots & (FR(Fa_n, Fb_1) \\ + & + & + \\ \vdots & \vdots & \vdots \\ + & + & + \\ FR(Fa_1, Fb_m))/m & \cdots & FR(Fa_n, Fb_m)) \,/\, m \end{bmatrix} = [\downarrow FR_1 \;\; \downarrow FR_2 \;\; \dots \;\; \downarrow FR_m]$$

$$SR\,(Ra, Rb) = \cos(\vec{M}, \downarrow M) = \frac{\vec{M} \,.\downarrow M}{\|\vec{M}\| \, \|\downarrow M\|}$$

Evaluation Plan



I: SEM-REQ Dataset

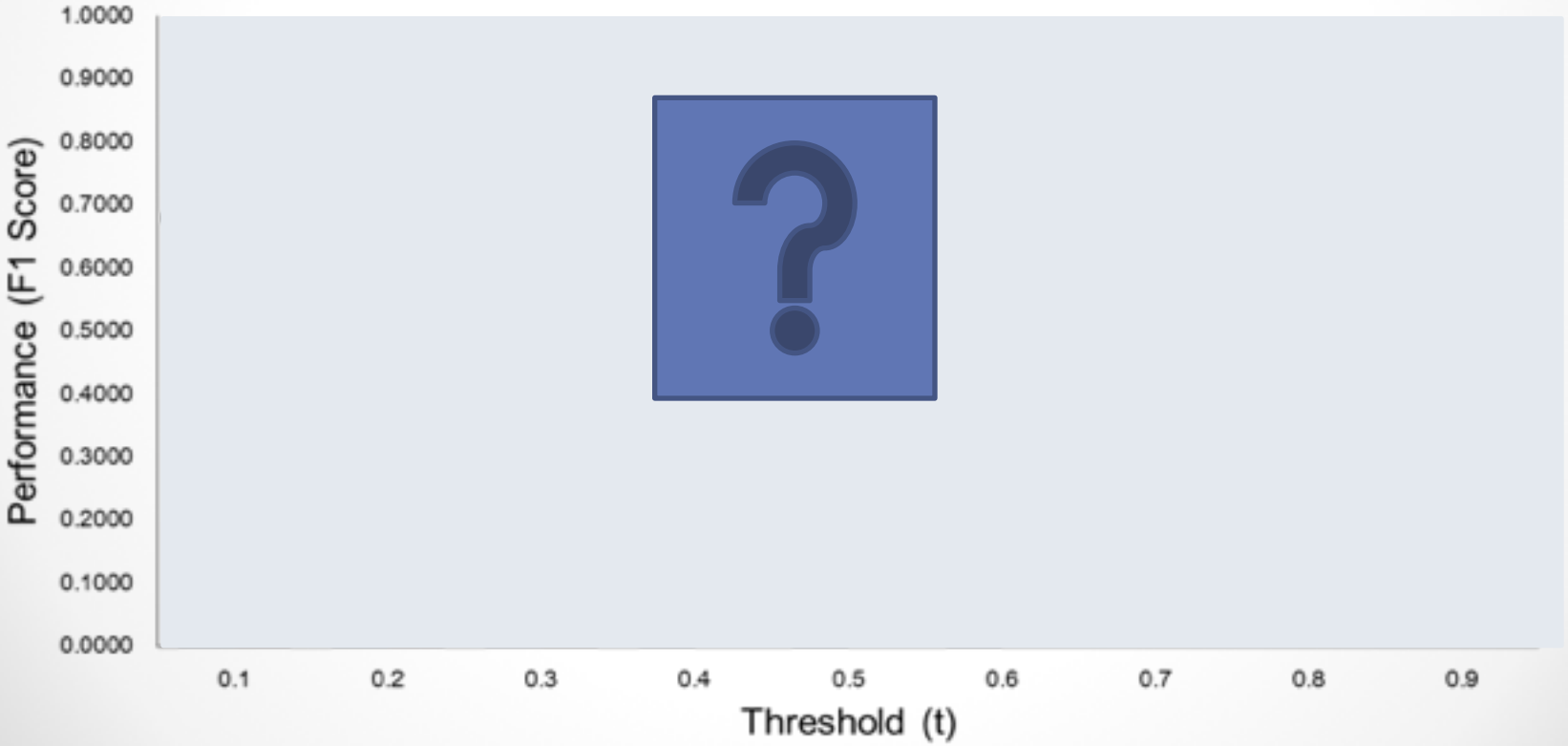
1. **Annotating 1770** requirements pairs from FN-RE corpus [2][3] by **3** annotators independently.
2. **Validating** the dataset
    1. with an average F-score of 77.5%
3. **Harmonisation** to produce final dataset of SEM-REQ.

II: Baseline System

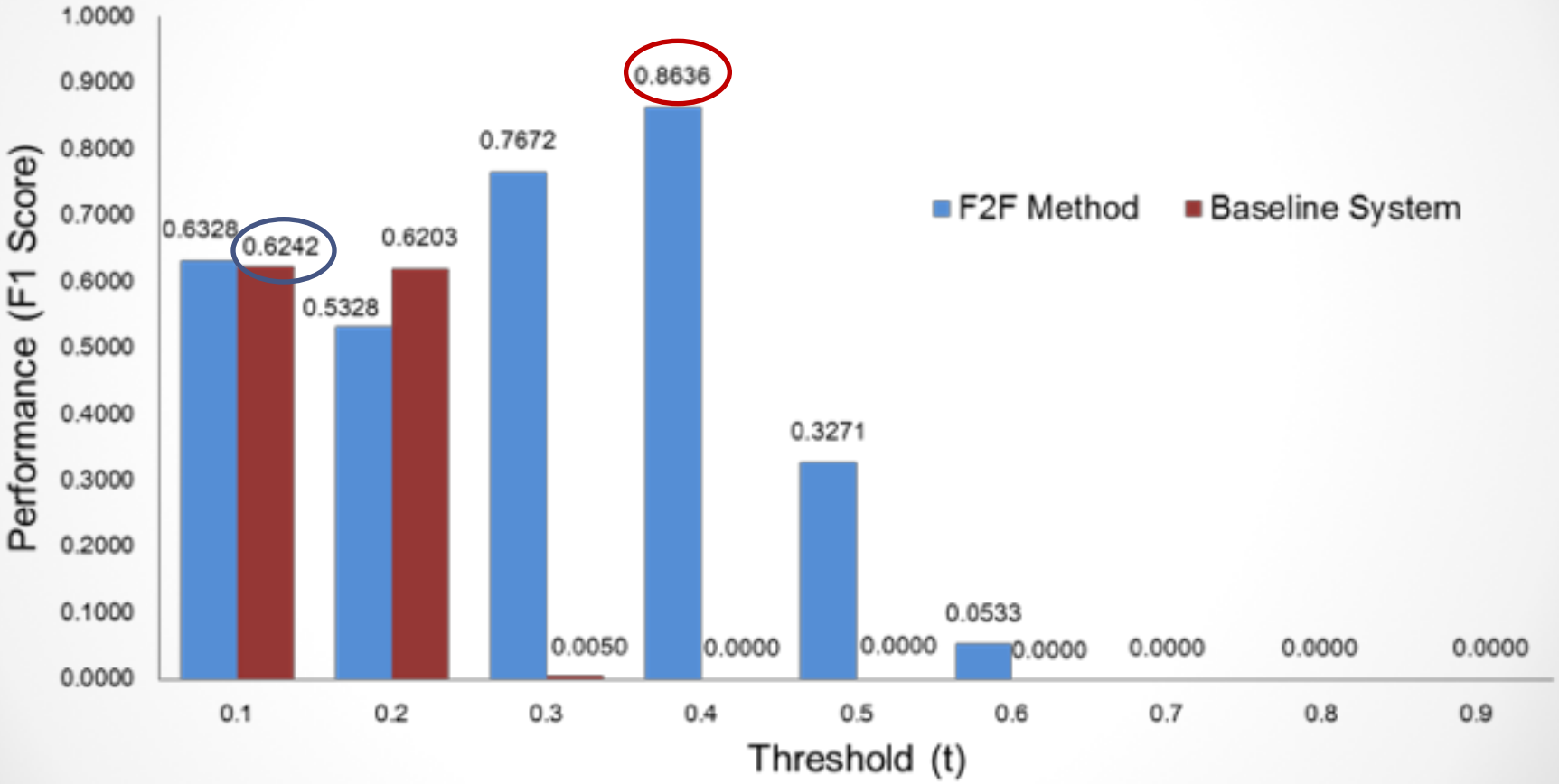
- Using by pre-trained word embeddings, i.e., Google's Word2Vec model.
- Applying same procedure of measuring semantic relatedness(i.e. cosine metric and embeddings averaging)

Preparing the Results

- We compared the F2F method with the baseline system by applying each of them to SEM-REQ.



F1 Scores Comparison



Examples

ID_A	Sentence_A	ID_B	Sentence_B	F2F Score
FN-REQ-005-2	He ACCESSES$_{Having_or_lacking_access}$ the website , CREATES$_{Creating}$ a profile and PROVIDES$_{Supply}$ his educational professional and personal information	FN-REQ-007-2	On registration , they NEED$_{Have_as_requirement}$ to PROVIDE$_{Supply}$ name and address , payment details (credit card , etc) , shoe sizes , gender , and any special details	0.528667032
FN-REQ-007-2	On registration , they NEED$_{Have_as_requirement}$ to PROVIDE$_{Supply}$ name and address , payment details (credit card , etc) , shoe sizes , gender , and any special details	FN-REQ-022-3	WHEN$_{Temporal_collocation}$ all items have been CHOSEN$_{Choosing}$, the shopper PROVIDES$_{Supply}$ a delivery address .	0.420046491
FN-REQ-030-5	John INDICATES$_{Indicating}$ that he WISHES$_{Desiring}$ to WITHDRAW$_{Removing}$ $ 50 dollars .	FN-REQ-030-8	The ATM VERIFIES$_{Verification}$ that the amount may be WITHDRAWN$_{Removing}$ from his account.	0.581840709
FN-REQ-015-9	After the Account Manager APPROVES$_{Deny_or_grant_permission}$ the purchase , an authorisation signature MAY$_{Possibility}$ be REQUIRED$_{Have_as_requirement}$.	FN-REQ-022-1	The Pizza Ordering SYSTEM$_{Gizmo}$ ALLOWS$_{Preventing_or_letting}$ the user of a web browser to ORDER$_{Request_entity}$ pizza for home delivery .	0.400741491
FN-REQ-007-1	Customers will NEED$_{Have_as_requirement}$ to REGISTER$_{Recording}$ with the Odd Shoe Company to MAKE$_{Intentionally_create}$ orders .	FN-REQ-015-9	After the Account Manager APPROVES$_{Deny_or_grant_permission}$ the purchase , an authorisation signature MAY$_{Possibility}$ be REQUIRED$_{Have_as_requirement}$.	0.509795616

Ongoing Work

...

End

References

[1] Fabiano Dalpiaz, Alessio Ferrari, Xavier Franch, and Cristina Palomares. **Natural language processing for requirements engineering: The best is yet to come**. IEEE Software, 35(5):115–119, 2018.

[2] Waad Alhoshan, Riza Batista-Navarro, and Liping Zhao**. A framenet-based approach for annotating software requirements**. In Tiago Timponi Torrent, Lars Borin, and Collin F. Baker, editors, Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Paris, France, 2018. European Language Resources Association (ELRA).

[3] Waad Alhoshan, Riza Batista-Navarro, and Liping Zhao. **Towards a corpus of requirements documents enriched with semantic frame annotations**. In 2018 IEEE 26th International Requirements Engineering Conference (RE), pages 428–431, 2018.

[4] Waad Alhoshan, Liping Zhao, and Riza Batista-Navarro. **Using semantic frames to identify related textual requirements: An initial validation**. In Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM '18, pages 58:1–58:2, New York, NY, USA, 2018. ACM.

[5] Waad Alhoshan, Riza Batista-Navarro, and Liping Zhao. **Semantic Frame Embeddings for Detecting Relations between Software Requirements**. 13th International Conference on Computational Semantics (IWCS). Association for Computational Linguistics, 2019.

End

Thank you

Waad Alhoshan
PhD Candidate, Software & Systems Research Group
School of Computer Science, University of Manchester
Email : Waad.alhoshan@postgrad.manchester.ac.uk