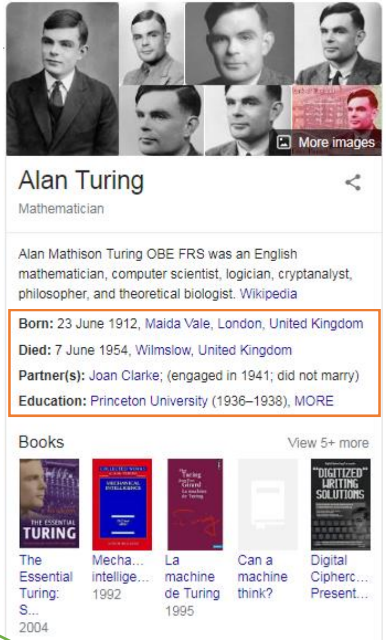
# DeepLENS: Deep Learning for Entity Summarization

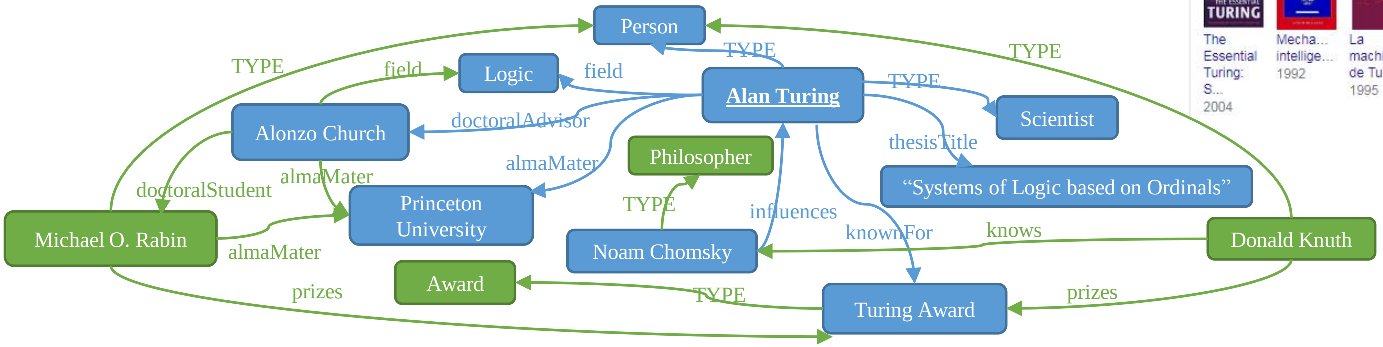
Qingxia Liu, Gong Cheng, and Yuzhong Qu

National Key Laboratory for Novel Software Technology, Nanjing University, China

# Entity Summarization (ES)

- RDF Graph: T
  - triple t∈T: <subj, pred, obj>
- Entity Description: Desc(e)
  - Desc(e) = {t∈T: subj(t)=e or obj(t)=e}
  - triple t∈Desc(e): <e, property, value>
- Entity Summarization: S(e, k)
  - S⊆Desc(e) , |S|≤k

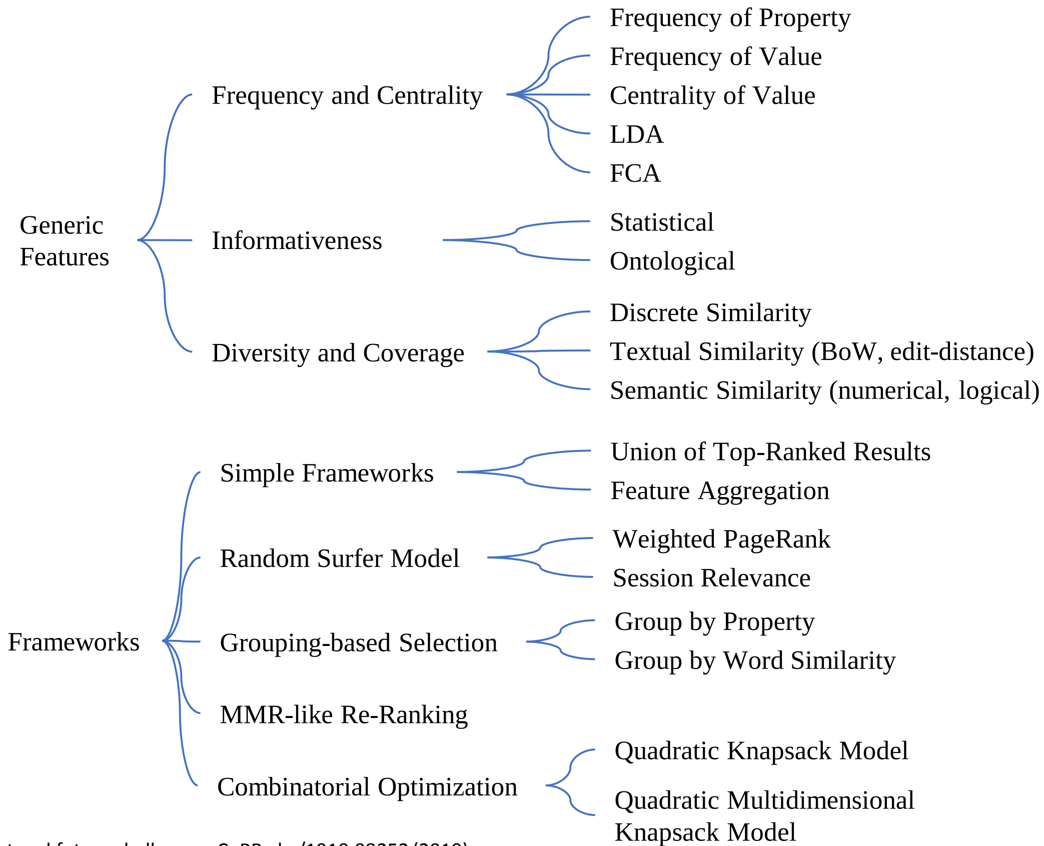


# Existing Solutions

## Un-Supervised Methods[1]

- RELIN
- DIVERSUM
- FACES
- FACES-E
- CD
- LinkSUM
- BAFREC
- KAFCA
- MPSUM
- ...




[1] Liu, Q., Cheng, G., Gunaratna, K., Qu, Y.: Entity summarization: state of the art and future challenges. CoRR abs/1910.08252 (2019)

# Existing Solutions

**Un-Supervised Methods**[1]

- RELIN
- DIVERSUM
- FACES
- FACES-E
- CD
- LinkSUM
- BAFREC
- KAFCA
- MPSUM
- ...

Can deep learning summarize better?

[1] Liu, Q., Cheng, G., Gunaratna, K., Qu, Y.: Entity summarization: state of the art and future challenges. CoRR abs/1910.08252 (2019)

# Existing Solutions

## Un-Supervised Methods

- RELIN
- DIVERSUM
- FACES
- FACES-E
- CD
- LinkSUM
- BAFREC
- KAFCA
- MPSUM
- ...

## Supervised Methods

- ESA[12]
  - graph embedding (TransE), BiLSTM

| | DBpedia | | LinkedMDB | | ALL | |
|---|---|---|---|---|---|---|
| | k=5 | k=10 | k=5 | k=10 | k=5 | k=10 |
| **RELIN** [4] | 0.242 | 0.455 | 0.203 | 0.258 | 0.231 | 0.399 |
| **DIVERSUM** [13] | 0.249 | 0.507 | 0.207 | 0.358 | 0.237 | 0.464 |
| **CD** [12] | 0.287 | 0.517 | 0.211 | 0.328 | 0.252 | 0.455 |
| **FACES−E** [7] | 0.280 | 0.485 | 0.313 | 0.393 | 0.289 | 0.461 |
| **FACES** [8] | 0.270 | 0.428 | 0.169 | 0.263 | 0.241 | 0.381 |
| **LinkSUM** [14] | 0.274 | 0.479 | 0.140 | 0.279 | 0.236 | 0.421 |
| **ESA** | **0.310** | **0.525** | **0.320** | **0.403** | **0.312** | **0.491** |

Table 1: Comparison of F-measure on ESBM benchmark v1.1

[12] Wei, D., Liu, Y., Zhu, F., Zang, L., Zhou, W., Han, J., Hu, S.: ESA: Entity summarization with attention. In: EYRE 2019. pp. 40-44 (2019)

# Existing Solutions

## Un-Supervised Methods

- RELIN
- DIVERSUM
- FACES
- FACES-E
- CD
- LinkSUM
- BAFREC
- KAFCA
- MPSUM
- ...

## Supervised Methods

- ESA[12]
  - graph embedding (TransE), BiLSTM

| | DBpedia | | LinkedMDB | | ALL | |
|---|---|---|---|---|---|---|
| | k=5 | k=10 | k=5 | k=10 | k=5 | k=10 |
| **RELIN** [4] | 0.242 | 0.455 | 0.203 | 0.258 | 0.231 | 0.399 |
| **DIVERSUM** [13] | 0.249 | 0.507 | 0.207 | 0.358 | 0.237 | 0.464 |
| **CD** [12] | 0.287 | 0.517 | 0.211 | 0.328 | 0.252 | 0.455 |
| **FACES–E** [7] | 0.280 | 0.485 | 0.313 | 0.393 | 0.289 | 0.461 |
| **FACES** [8] | 0.270 | 0.428 | 0.169 | 0.263 | 0.241 | 0.381 |
| **LinkSUM** [14] | 0.274 | 0.479 | 0.140 | 0.279 | 0.236 | 0.421 |
| **ESA** | **0.310** | **0.525** | **0.320** | **0.403** | **0.312** | **0.491** |

Table 1: Comparison of F-measure on ESBM benchmark v1.1

small improvement

- +7% compared with unsupervised FACES-E

[12] Wei, D., Liu, Y., Zhu, F., Zang, L., Zhou, W., Han, J., Hu, S.: ESA: Entity summarization with attention. In: EYRE 2019. pp. 40-44 (2019)

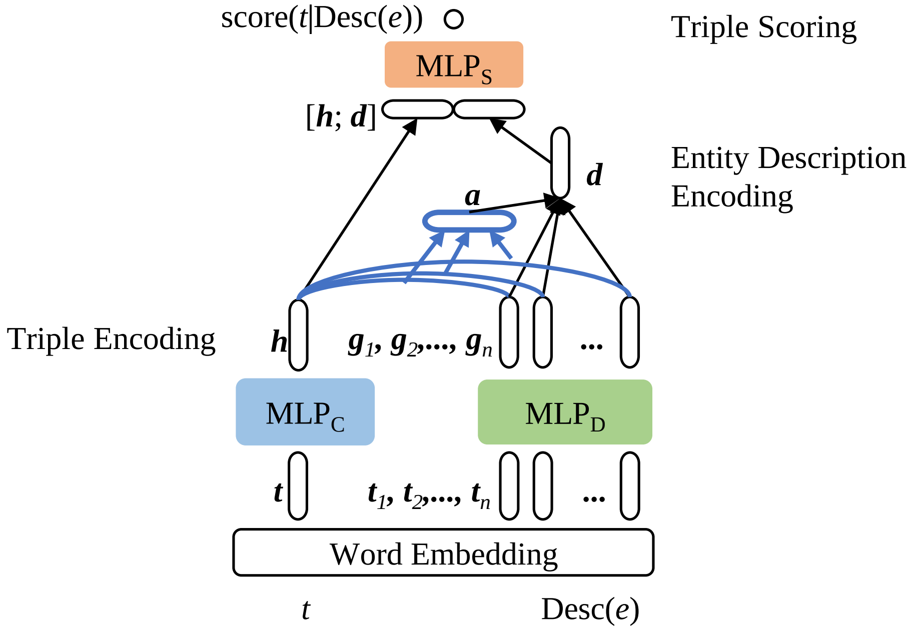
# Our Idea

- Deisgn a novel Deep Learning based approach to Entity Summarization: **DeepLENS**
  - Entity summary presented as short **text**:  textual semantics
  - Entity description as a triple **set**:  permutation invariant

| | **ESA** | **DeepLENS** |
|---|---|---|
| Triple Encoding | Graph Embedding | Word Embedding |
| Triple Set Encoding | Sequence Model | Aggregation-based Model |

# Our Solution

- DeepLENS
  - Triple Encoding
  - Entity Description Encoding
  - Triple Scoring

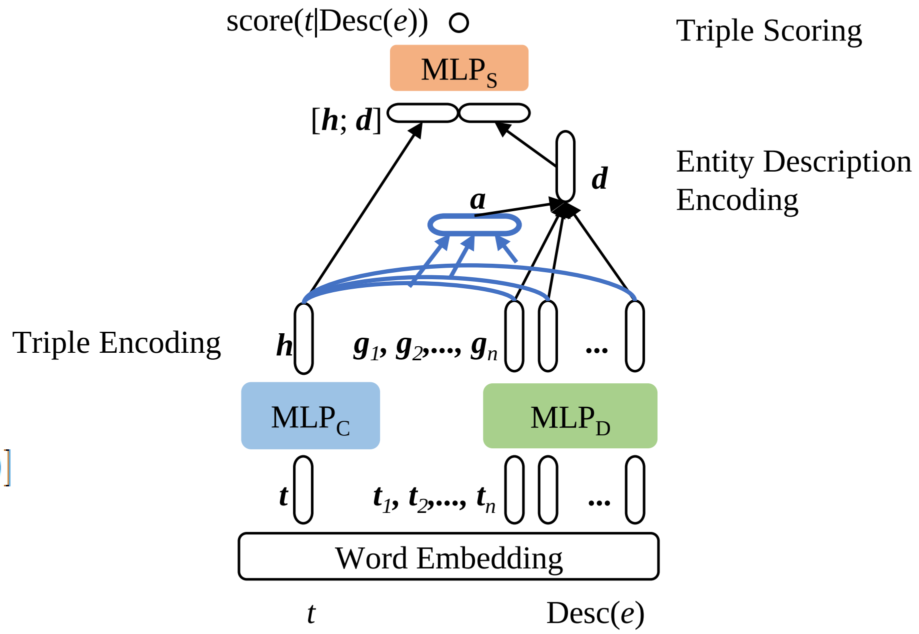


# Our Solution

- DeepLENS
  - *Triple Encoding*
  - Entity Description Encoding
  - Triple Scoring

textual semantics of triple

$$t = [\mathrm{Embedding}(\texttt{prop}(t));\ \mathrm{Embedding}(\texttt{val}(t))]$$
$$h = \mathrm{MLP}_{\mathrm{C}}(t)$$



# Our Solution

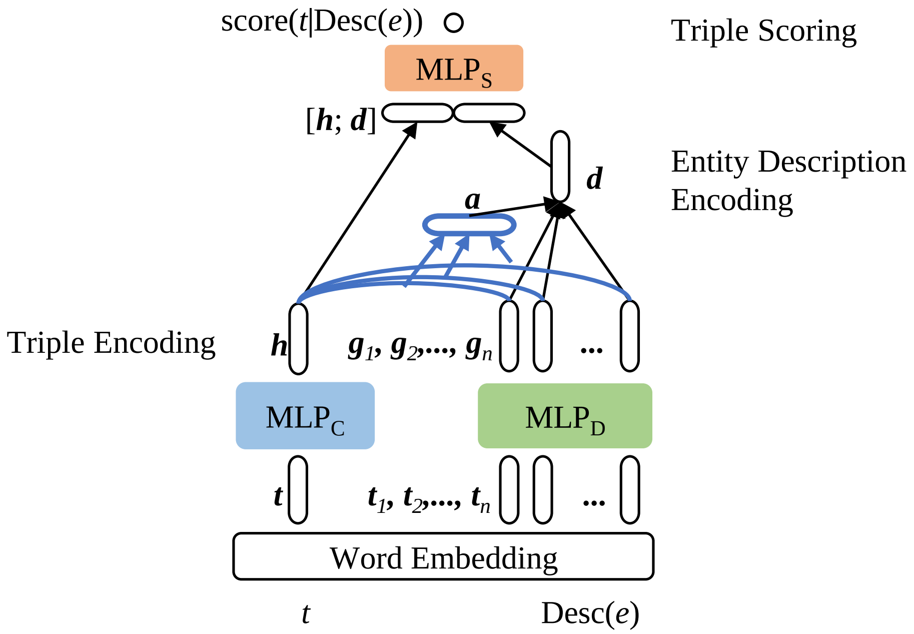
- DeepLENS
  - Triple Encoding
  - *Entity Description Encoding*
  - Triple Scoring

permutation invariant representation

$$g_i = \mathrm{MLP_D}(t_i)$$

$$a_i = \frac{\exp(\cos(h, g_i))}{\sum_j \exp(\cos(h, g_j))}$$

$$d = \sum_{i=1}^{n} a_i g_i$$

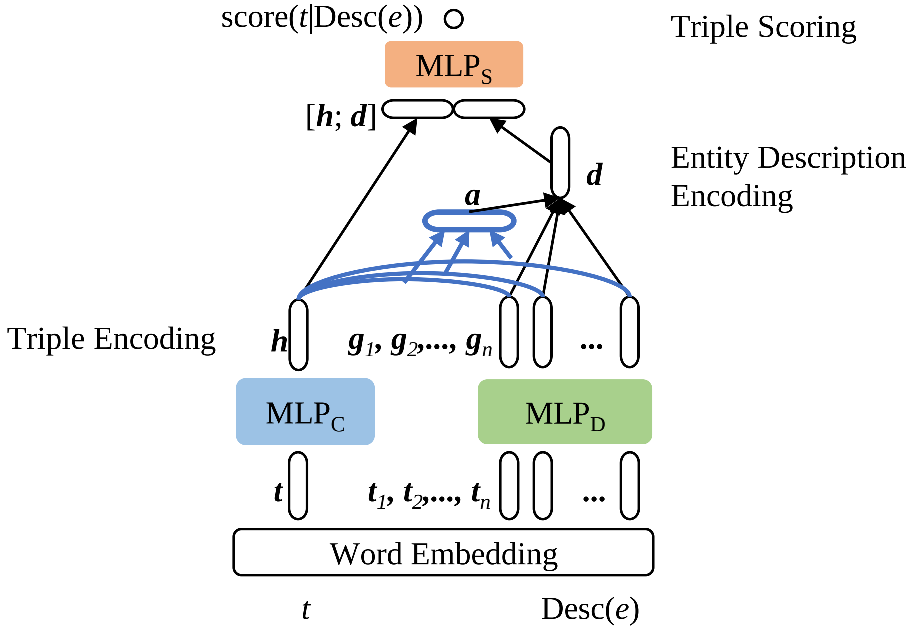


# Our Solution

- DeepLENS
  - Triple Encoding
  - Entity Description Encoding
  - *Triple Scoring*

context-based salience score

$$\texttt{score}(t|\texttt{Desc}(e)) = \mathrm{MLP_S}([\boldsymbol{h};\ \boldsymbol{d}])$$



# Evaluation

- Dataset: ESBM v1.2[8]
  - DBpedia, LinkedMDB
- Metric: F1
- Participating Methods
  - Unsupervised Methods
    - RELIN, DIVERSUM, FACES, FACES-E, CD, LinkSUM, BAFREC, KAFCA, MPSUM
  - Supervised Methods
    - ESA (state of the art)
    - DeepLENS (our method)
  - Oracle Method
    - ORACLE (best possible performance on ESBM)
      - Summary consisting of k triples that most frequently appear in ground-truth summaries

[8] Liu, Q., Cheng, G., Gunaratna, K., Qu, Y.: ESBM: An entity summarization benchmark. In: ESWC 2020 (2020)

# Overall Result

- Results
  - supervised > unsupervised
  - DeepLENS > all baselines
  - ORACLE > DeepLENS
    - suggesting room for improvement

**Table 1.** Average F1 over all the test entities. Significant and insignificant differences ($p < 0.01$) between DeepLENS and each baseline are indicated by ▲ and ∘, respectively.

| | DBpedia | | LinkedMDB | |
|---|---|---|---|---|
| | $k = 5$ | $k = 10$ | $k = 5$ | $k = 10$ |
| RELIN [2] | 0.242 | 0.455 | 0.203 | 0.258 |
| DIVERSUM [9] | 0.249 | 0.507 | 0.207 | 0.358 |
| FACES [3] | 0.270 | 0.428 | 0.169 | 0.263 |
| FACES-E [4] | 0.280 | 0.488 | 0.313 | 0.393 |
| CD [13] | 0.283 | 0.513 | 0.217 | 0.331 |
| LinkSUM [10] | 0.287 | 0.486 | 0.140 | 0.279 |
| BAFREC [6] | 0.335 | 0.503 | 0.360 | 0.402 |
| KAFCA [5] | 0.314 | 0.509 | 0.244 | 0.397 |
| MPSUM [11] | 0.314 | 0.512 | 0.272 | 0.423 |
| ESA [12] | 0.331 | 0.532 | 0.350 | 0.416 |
| DeepLENS | 0.404 ▲▲▲▲▲▲▲▲▲▲ | 0.575 ▲▲▲▲▲▲▲▲▲▲ | 0.469 ▲▲▲▲▲▲▲▲▲▲ | 0.489 ▲▲▲▲▲▲▲▲▲▲ |
| ORACLE | 0.595 | 0.713 | 0.619 | 0.678 |

# Conclusion

- Presented a simple yet effective deep learning model for ES.
  - textual semantics
  - permutation invariance
- Achieved new state-of-the-art results on the ESBM benchmark.
- ES can be effectively solved with properly designed deep learning models.

- Future Work
  - ontological semantics
  - structural semantics

# Main Conference Papers

- **Entity Summarization with User Feedback**

  Qingxia Liu, Yue Chen, Gong Cheng, Evgeny Kharlamov, Junyou Li and Yuzhong Qu

  - Session 3: Extraction and Recommendation 2
  - Thursday, June 4, 10:20-10:40

- **ESBM: An Entity Summarization Benchmark**

  Qingxia Liu, Gong Cheng, Kalpa Gunaratna and Yuzhong Qu

  - Session 9: Benchmarking
  - Thursday, June 4, 11:50-12:10

Thank you !

Questions ?