# Translating Embeddings for Modeling Multi-relational Data

Advisor: Jia-Ling, Koh

Speaker: Hsiang-Hui Chou

Date:2022/10/04

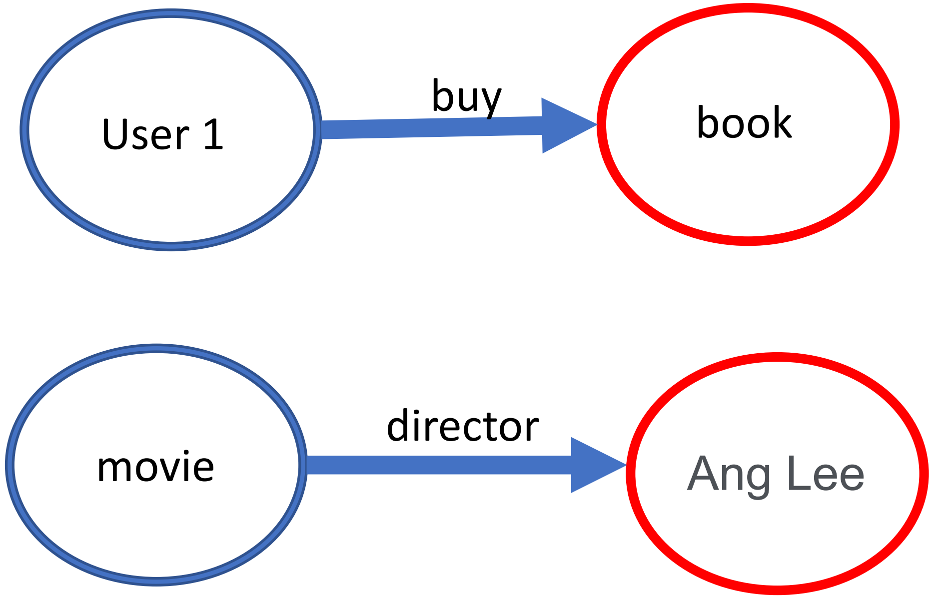
# Outline

- Introduction
- Method
- Experiment
- Conclusion

# Multi-relational Data

(head, label, tail) $(h, l, t)$

ex:
(User 1, buy, book)
(movie, director, Ang Lee)

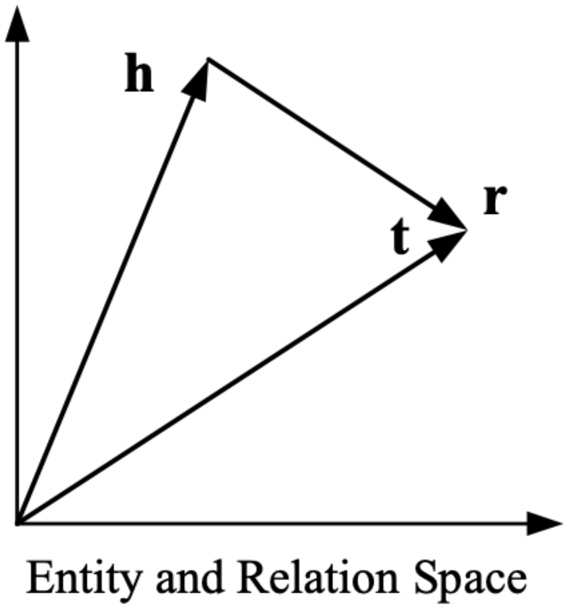



Node : entity
Edge : relationship

# Outline

- Introduction
- Method
- Experiment
- Conclusion

# Translation-based model

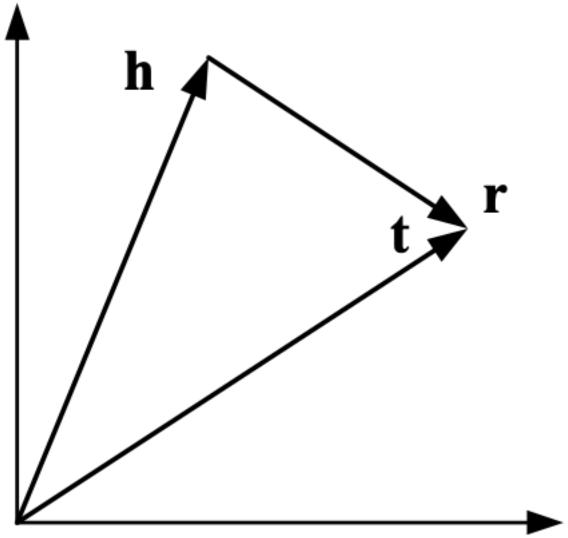



$$\mathcal{L} = \sum_{(h,\ell,t)\in S} \sum_{(h',\ell,t')\in S'_{(h,\ell,t)}} \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$$

Negative triple : head or tail replaced by a random entity

$$S'_{(h,\ell,t)} = \left\{(h',\ell,t) | h' \in E\right\} \cup \left\{(h,\ell,t') | t' \in E\right\}$$

# Translation-based model




Entity and Relation Space

$$d(\boldsymbol{h}+\boldsymbol{\ell},\boldsymbol{t}) = \|\boldsymbol{h}\|_2^2 + \|\boldsymbol{\ell}\|_2^2 + \|\boldsymbol{t}\|_2^2 - 2\big(\boldsymbol{h}^T\boldsymbol{t} + \boldsymbol{\ell}^T(\boldsymbol{t}-\boldsymbol{h})\big)\,.$$

$$\boldsymbol{h}^T\boldsymbol{t} + \boldsymbol{\ell}^T(\boldsymbol{t}-\boldsymbol{h})$$

**Algorithm 1** Learning TransE

**input** Training set $S = \{(h, \ell, t)\}$, entities and rel. sets $E$ and $L$, margin $\gamma$, embeddings dim. $k$.

**initialize** $\boldsymbol{\ell} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$
    $\boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$
    $\mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$
**loop**
  $\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$
  $S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size $b$
  $T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets
  **for** $(h, \ell, t) \in S_{batch}$ **do**
    $(h', \ell, t') \leftarrow \text{sample}(S'_{(h,\ell,t)})$ // sample a corrupted triplet
    $T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$
  **end for**
  Update embeddings w.r.t. $\displaystyle\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$
**end loop**

# Outline

- Introduction
- Method
- Experiment
- Conclusion

# Dataset

Table 2: **Statistics of the data sets** used in this paper and extracted from the two knowledge bases, Wordnet and Freebase.

| DATA SET | WN | FB15K | FB1M |
|---|---|---|---|
| ENTITIES | 40,943 | 14,951 | $1\times10^6$ |
| RELATIONSHIPS | 18 | 1,345 | 23,382 |
| TRAIN. EX. | 141,442 | 483,142 | $17.5\times10^6$ |
| VALID EX. | 5,000 | 50,000 | 50,000 |
| TEST EX. | 5,000 | 59,071 | 177,404 |

# Dataset-Wordnet

| Semantic Relation | Syntactic Category | Examples |
|---|---|---|
| Synonymy (similar) | N, V, Aj, Av | pipe, tube<br>rise, ascend<br>sad, unhappy<br>rapidly, speedily |
| Antonymy (opposite) | Aj, Av, (N, V) | wet, dry<br>powerful, powerless<br>friendly, unfriendly<br>rapidly, slowly |
| Hyponymy (subordinate) | N | sugar maple, maple<br>maple, tree<br>tree, plant |
| Meronymy (part) | N | brim, hat<br>gin, martini<br>ship, fleet |
| Troponomy (manner) | V | march, walk<br>whisper, speak |
| Entailment | V | drive, ride<br>divorce, marry |

***Note:*** ***N*** *= Nouns* ***Aj*** *= Adjectives* ***V*** *= Verbs* ***Av*** *= Adverbs*

# Dataset -Freebase

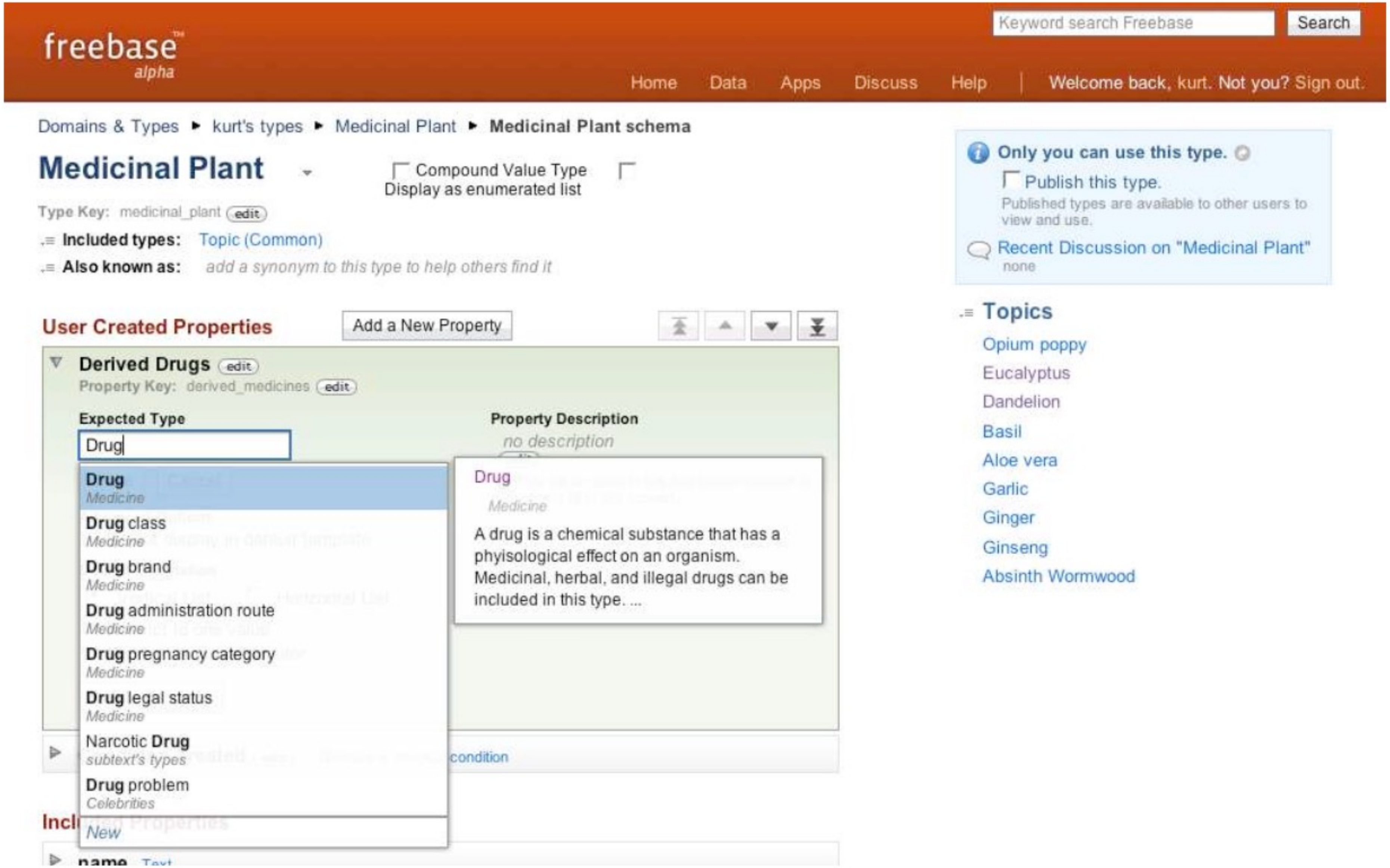


Table 1: **Numbers of parameters** and their values for FB15k (in millions). $n_e$ and $n_r$ are the nb. of entities and relationships; $k$ the embeddings dimension.

| METHOD | NB. OF PARAMETERS | ON FB15K |
|---|---|---|
| Unstructured [2] | $O(n_e k)$ | 0.75 |
| RESCAL [11] | $O(n_e k + n_r k^2)$ | 87.80 |
| SE [3] | $O(n_e k + 2n_r k^2)$ | 7.47 |
| SME(LINEAR) [2] | $O(n_e k + n_r k + 4k^2)$ | 0.82 |
| SME(BILINEAR) [2] | $O(n_e k + n_r k + 2k^3)$ | 1.06 |
| LFM [6] | $O(n_e k + n_r k + 10k^2)$ | 0.84 |
| TransE | $O(n_e k + n_r k)$ | 0.81 |

Table 3: **Link prediction results.** Test performance of the different methods.

| DATASET | WN | | | | FB15K | | | | FB1M | |
|---|---|---|---|---|---|---|---|---|---|---|
| METRIC | MEAN RANK | | HITS@10 (%) | | MEAN RANK | | HITS@10 (%) | | MEAN RANK | HITS@10 (%) |
| *Eval. setting* | *Raw* | *Filt.* | *Raw* | *Filt.* | *Raw* | *Filt.* | *Raw* | *Filt.* | *Raw* | *Raw* |
| Unstructured [2] | 315 | 304 | 35.3 | 38.2 | 1,074 | 979 | 4.5 | 6.3 | 15,139 | 2.9 |
| RESCAL [11] | 1,180 | 1,163 | 37.2 | 52.8 | 828 | 683 | 28.4 | 44.1 | - | - |
| SE [3] | 1,011 | 985 | 68.5 | 80.5 | 273 | 162 | 28.8 | 39.8 | 22,044 | 17.5 |
| SME(LINEAR) [2] | 545 | 533 | 65.1 | 74.1 | 274 | 154 | 30.7 | 40.8 | - | - |
| SME(BILINEAR) [2] | 526 | 509 | 54.7 | 61.3 | 284 | 158 | 31.3 | 41.3 | - | - |
| LFM [6] | 469 | 456 | 71.4 | 81.6 | 283 | 164 | 26.0 | 33.1 | - | - |
| TransE | **263** | **251** | **75.4** | **89.2** | **243** | **125** | **34.9** | **47.1** | **14,615** | **34.0** |

Table 4: **Detailed results by category of relationship.** We compare Hits@10 (in %) on FB15k in the filtered evaluation setting for our model, TransE and baselines. (M. stands for MANY).

| TASK | PREDICTING *head* | | | | PREDICTING *tail* | | | |
|---|---|---|---|---|---|---|---|---|
| REL. CATEGORY | 1-TO-1 | 1-TO-M. | M.-TO-1 | M.-TO-M. | 1-TO-1 | 1-TO-M. | M.-TO-1 | M.-TO-M. |
| Unstructured [2] | 34.5 | 2.5 | 6.1 | 6.6 | 34.3 | 4.2 | 1.9 | 6.6 |
| SE [3] | 35.6 | 62.6 | 17.2 | 37.5 | 34.9 | 14.6 | 68.3 | 41.3 |
| SME(LINEAR) [2] | 35.1 | 53.7 | 19.0 | 40.3 | 32.7 | 14.9 | 61.6 | 43.3 |
| SME(BILINEAR) [2] | 30.9 | **69.6** | **19.9** | 38.6 | 28.2 | 13.1 | **76.0** | 41.8 |
| TransE | **43.7** | 65.7 | 18.2 | **47.2** | **43.7** | **19.7** | 66.7 | **50.0** |

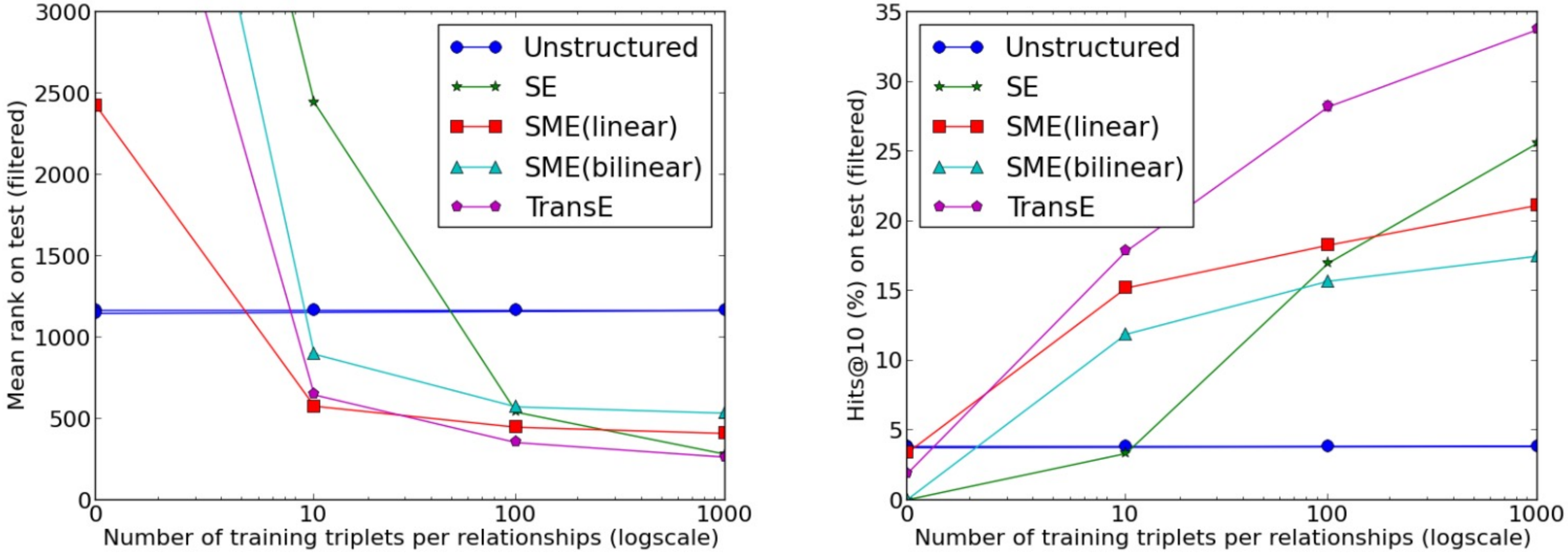



Figure 1: **Learning new relationships with few examples**. Comparative experiments on FB15k data evaluated in mean rank (left) and hits@10 (right). More details in the text.

Table 5: **Example predictions** on the FB15k test set using TransE. **Bold** indicates the test triplet's true tail and *italics* other true tails present in the training set.

| INPUT (HEAD AND LABEL) | PREDICTED TAILS |
|---|---|
| J. K. Rowling `influenced by` | *G. K. Chesterton*, J. R. R. Tolkien, *C. S. Lewis*, **Lloyd Alexander**, Terry Pratchett, Roald Dahl, Jorge Luis Borges, *Stephen King*, Ian Fleming |
| Anthony LaPaglia `performed in` | *Lantana*, *Summer of Sam*, *Happy Feet*, *The House of Mirth*, Unfaithful, **Legend of the Guardians**, Naked Lunch, X-Men, The Namesake |
| Camden County `adjoins` | **Burlington County**, *Atlantic County*, *Gloucester County*, Union County, Essex County, New Jersey, Passaic County, Ocean County, Bucks County |
| The 40-Year-Old Virgin `nominated for` | *MTV Movie Award for Best Comedic Performance,* *BFCA Critics' Choice Award for Best Comedy,* *MTV Movie Award for Best On-Screen Duo,* MTV Movie Award for Best Breakthrough Performance, **MTV Movie Award for Best Movie**, MTV Movie Award for Best Kiss, D. F. Zanuck Producer of the Year Award in Theatrical Motion Pictures, Screen Actors Guild Award for Best Actor - Motion Picture |
| Costa Rica football team `has position` | *Forward*, *Defender*, *Midfielder*, **Goalkeepers**, Pitchers, Infielder, Outfielder, Center, Defenseman |
| Lil Wayne `born in` | **New Orleans**, Atlanta, Austin, St. Louis, Toronto, New York City, Wellington, Dallas, Puerto Rico |
| WALL-E `has the genre` | Animations, Computer Animation, *Comedy film*, *Adventure film*, *Science Fiction*, **Fantasy**, Stop motion, *Satire*, Drama |

# Outline

- Introduction
- Method
- Experiment
- Conclusion

# Conclusion

- proposed a new approach to learn embeddings of KBs, focusing on the minimal parametrization
- highly scalable model