UOM-FJ at the NTCIR-15 SHINRA2020-ML Task

Hiyori Yoshikawa¹ Chunpeng Ma¹ Aili Shen² Qian Sun²
Chenbang Huang² Guillaume Pelat³ Akiba Miura¹ Daniel Beck²
Timothy Baldwin² Tomoya Iwakura¹

¹Fujitsu Laboratories Ltd., Japan ²University of Melbourne, Australia
³École Polytechnique, France
We participated in 28 subtasks of the SHINRA2020-ML Task: mapping Wikipedia entities into Extended Named Entity (ENE) categories.

Our model was trained to capture multiple aspects of Wikipedia articles: text, structured knowledge, images, page layout and the ENE class hierarchy.

Our system ranked first in four languages, achieving an F1-score of 82.73 on the English subtask.
Our System Overview

Document representation:  \[ e = \text{FFNN}(e_{\text{BERT}} \oplus e_{\text{KG}} \oplus e_{\text{VL}} \oplus e_{\text{SS}} \oplus e_{\text{CH}}) \]

- BERT encoder (Devlin et al., 2019) for **pure textual information** \( e_{\text{BERT}} \)

- A set of separately-trained document representations that encodes different aspects of a given document:
  - knowledge graph features \( e_{\text{KG}} \), text and images \( e_{\text{VL}} \),
  - page screenshot layout \( e_{\text{SS}} \), and ENE class hierarchy \( e_{\text{CH}} \)

Label probability:  \[ p(c, e) = \sigma(w_{c}^T e + b_{c}) \]
Knowledge graph ($e_{KG}$)

- Pre-trained embeddings of Wikidata graph (Lerer et al., 2019)
- `wikibase_item` field of the Wikipedia dump is used to identify the Wikidata entity corresponding to a Wikipedia article
- 98.3% of English Wikipedia articles have corresponding Wikidata entries

https://www.wikidata.org/wiki/Q1490
Text–image representation \( (e_{VL}) \)

- **VL-BERT**-based (Su et al., 2020) encoding of text and images in a given Wikipedia article
- Multiple images are rescaled and concatenated to compose a single large image; individual images are treated as regions-of-interest (ROIs)
- The VL-BERT model is fine-tuned on the SHINRA2020-ML task

---

**Tokyo** \( (\text{toʊkiоu} \ TOH-këe-oh, \text{-kjou} -\text{k}yoh; \text{Japanese: 東京, Tōkyō [toːkio] (listen))}, \text{officially Tokyo Metropolis (東京都, Tokyo-to), is the capital and most populous prefecture of Japan. Located at the head of Tokyo Bay, the prefecture forms part of the Kantō region on the central Pacific coast of Japan’s main island of Honshu. ...} \)
Page screenshot layout ($e_{SS}$)

- We obtain visual renderings of Wikipedia articles from a Wikipedia dump to generate their screenshots.

- The screenshots are resized to a fixed size and fed into INCEPTION (Inception V3 with different convolution filters) (Szegedy et al., 2016) to obtain visual representations.
ENE class hierarchy \( (e_{CH}) \)

- We employ a hierarchy-aware global model (HiGAM) (Zhou et al., 2020) to capture the label hierarchy and correlations between ENE categories.
- Construct hierarchy graph: node = ENE category, edge = hierarchical relational information or correlation information.
- The encoder is trained with binary cross-entropy loss over the hierarchical label space.

http://ene-project.info/ene8/?lang=en
We examined three ways to generate the final outputs:

- jointrep: output the categories whose estimated probability exceeds the threshold $\theta = 0.5$

- jointrepPostprocess: jointrep + assign the label CONCEPT if the probability scores of all candidate categories are below the threshold

- jointrepUnionPostprocess: take the union of different predictions made by five models trained on different subsets of the training data, and then apply the post-processing step as in jointrepPostprocess
Main Results

Official evaluation results (English)

<table>
<thead>
<tr>
<th>Submission name</th>
<th>$P$</th>
<th>$R$</th>
<th>$F_1$</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>jointrep</td>
<td>81.77</td>
<td>83.71</td>
<td>82.73</td>
<td>1</td>
</tr>
<tr>
<td>jointrepPostprocess</td>
<td>81.46</td>
<td>83.71</td>
<td>82.57</td>
<td>3</td>
</tr>
<tr>
<td>jointrepUnionPostprocess</td>
<td>80.66</td>
<td>84.80</td>
<td>82.68</td>
<td>2</td>
</tr>
<tr>
<td>Other teams best</td>
<td>79.65</td>
<td>85.00</td>
<td>82.23</td>
<td></td>
</tr>
</tbody>
</table>

Other languages

- Non-English results are based on VL-BERT model, i.e. using only text–image representation
- Ranked first on Spanish, Italy and Catalan as well
- Relatively low performance on Arabic, Hindi and Thai
Ablation Study

leaderboard results, jointrep, English

<table>
<thead>
<tr>
<th>Model</th>
<th>$\mathcal{P}$</th>
<th>$\mathcal{R}$</th>
<th>$\mathcal{F}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>75.3</td>
<td>76.4</td>
<td>75.7</td>
</tr>
<tr>
<td>– knowledge graph</td>
<td>74.8</td>
<td>75.8</td>
<td>75.2</td>
</tr>
<tr>
<td>– text–image</td>
<td>74.5</td>
<td>75.5</td>
<td>74.8</td>
</tr>
<tr>
<td>– page layout</td>
<td>75.5</td>
<td>77.2</td>
<td>76.0</td>
</tr>
<tr>
<td>– class hierarchy</td>
<td>74.4</td>
<td>76.4</td>
<td>75.1</td>
</tr>
<tr>
<td>BERT + knowledge graph</td>
<td>70.8</td>
<td>71.3</td>
<td>71.0</td>
</tr>
<tr>
<td>BERT + text–image</td>
<td>73.5</td>
<td>74.8</td>
<td>73.9</td>
</tr>
<tr>
<td>BERT + page layout</td>
<td>70.5</td>
<td>71.6</td>
<td>70.9</td>
</tr>
<tr>
<td>BERT + class hierarchy</td>
<td>74.2</td>
<td>75.2</td>
<td>74.5</td>
</tr>
<tr>
<td>Page ID</td>
<td>Title</td>
<td>Opening text</td>
<td>Predicted ENES</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>54866853</td>
<td>Hedi (Policy)</td>
<td>Hedi was an economic policy of the imperial China. It was the state purchase of food supplies from farmers. As a means to control the price of grains and foods, it is an early example of Government procurement. The policy was adopted in the year of 488 by Emperor Xiaowen of Northern Wei as a counter measure of drought. The state purchases food supplies and stock them. ...</td>
<td>1.7.21.8:Plan</td>
</tr>
<tr>
<td>299259</td>
<td>Roman currency</td>
<td>Roman currency for most of Roman history consisted of gold, silver, bronze, orichalcum and copper coinage (see Roman metallurgy). From its introduction to the Republic, during Imperial times, Roman currency saw imination, and composition. A persistent debasement and replacement of coins over the centuries. ...</td>
<td>1.7.25.1:Currency</td>
</tr>
<tr>
<td>788538</td>
<td>Azamino</td>
<td>Azamino (あさみ野) is a bedroom community of Tokyo and Yokohama, located in Aoba-ku, Yokohama, Kanagawa Prefecture, Japan. The area is 20 minutes from Shibuya Station on the Tokyu Den-en-toshi Line and 27 minutes by car. Known as a refer to the area in the years. It is located next to the intense rama plaza.</td>
<td>1.5.0:Location_.Other</td>
</tr>
<tr>
<td>22786903</td>
<td>Hiro H1H</td>
<td>The Hiro H1H (or Navy Type 15) was a 1920s Japanese bomber or reconnaissance biplane flying boat developed from the Felixstowe F.5 by the Hiro Naval Arsenal for the Imperial Japanese Navy. The aircraft were built by Hiro, the Yokosuka Naval Arsenal and Aichi.</td>
<td>1.7.6:Weapon, 1.7.17.3:Aircraft</td>
</tr>
<tr>
<td>586540</td>
<td>Tailor</td>
<td>A tailor is a person who makes, repairs, or alters clothing professionally, especially suits and men's clothing. Although the term dates to the thirteenth century, and now refers to makers of men's and women's suits, coats, trousers, and similar garments, usually of wool, linen, or silk. ...</td>
<td>1.7.23.1:Position_:Vocation</td>
</tr>
</tbody>
</table>

**Redirect from Tailoring**

**Page redirect to different entity**

Tailoring
We participated in 28 subtasks of the SHINRA2020-ML Task: mapping Wikipedia entities into Extended Named Entity (ENE) categories.

Our model was trained to capture multiple aspects of Wikipedia articles: text, structured knowledge, images, page layout and the ENE class hierarchy.

Our system ranked first in four languages, achieving an F1-score of 82.73 on the English subtask.
References


