

Supplemental Material for: “Zero-Shot Open Entity Typing as Type-Compatible Grounding”

A Coverage Analysis for Concept Selection

To better understand the behavior of our concept retrieval, we perform an upper-bound analysis. Assuming that our type inference given concepts is perfect (an oracle inference), we want to estimate the best achievable score (upper-bound analysis). We perform this analysis in the output of two stages in the system,

1. Step 1: After initial concepts extraction (§3.1).
2. Step 2: After ELMO reranking (§3.2).

Figure 4 shows a summary of the upper-bound curves for the output of the Step 1 and the Step 2. From these analysis we set the parameters ℓ_{ESA} and ℓ_{ELMO} . From the blue curve it is evident that after about 200 concepts, there is almost high-coverage. With this, we set the parameter $\ell_{\text{ESA}} = 300$. Furthermore, the red curve shows the strong coverage, even with a few dozen candidates. In our experiments, we choose top $\ell_{\text{ELMO}} = 20$ concepts in the output of Step 2.

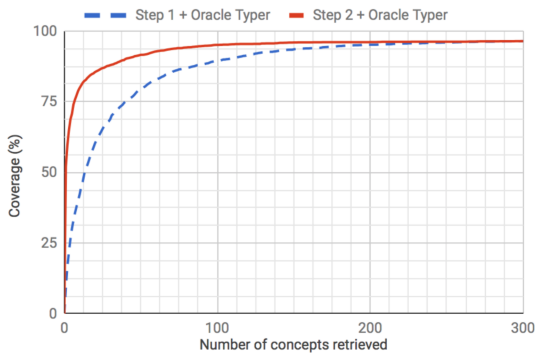


Figure 4: Upper-bound analysis: The y-axis shows the percentage of the instances with their types covered by the retrieved concepts. The x-axis shows the number of concepts retrieved, with a) Step 1 (dashed), b) Step 2 (solid).

B Output for Each Step: An Example

To give a better intuition about every single step in our algorithm, we plot the outputs of each step in our construction in Figure 5. The details of how each step is done is included in §3.

C Error Analysis

We expand a little bit on the errors analysis introduced in §4.6. The first two categories consist of the scenarios when context and mention-surface information result in a type-compatible concept (hence incorrect type). The last two categories are due to inaccurate mapping of (mostly type-compatible) concepts to a collection of types.

1. *Inconsistent concept, due to context information:* A short and ambiguous context could result in noisy concepts. In the following example, the majority of the selected concepts based on context are of type `/politician`, however the correct label is `/event`:

[The Fellows Forum], concerned in part with the induction of newly elected fellows, is just one event of the associations annual meeting.

2. *Inconsistent concept, due to surface information:* When a mention used in a sense other than its most-popular sense, it could result in mistakes. In the following example, while from the context it is clear that “Utah” is a `sports_team`, the surface-string has a stronger association with `Colorado`, which incorrectly results in the type `/location`:

The biggest cause for concern for McGuff is the bruised hamstring Regina Rogers suffered against [Utah] last Saturday.

3. *Incorrect type, due to type inference:* even when the system is able to find type-compatible concepts, it still fails to infer the correct type, if the types attached to the type-compatible are the majority among other types. In the following example, while there are concepts of type `/person`, the overall decision is incorrectly dominated by other concepts of type `/location`:

After years of wanting to curate such an exhibition, Wieczorek has collaborated with the Henry Art Gallery to feature 26 pieces of [Balzh]’s “Videowatercolors.”

The following is another example where the system fails to find the correct fine-type `/person/musician`, since the overall decision is dominated by other concepts with type `/person`, but not a musician (e.g. `Blair.Waldorf`).

“That’s what I love about [Balzh]’s art.”

4. *Incorrect type, due to type definition:* Some errors are caused by inaccurate definition for

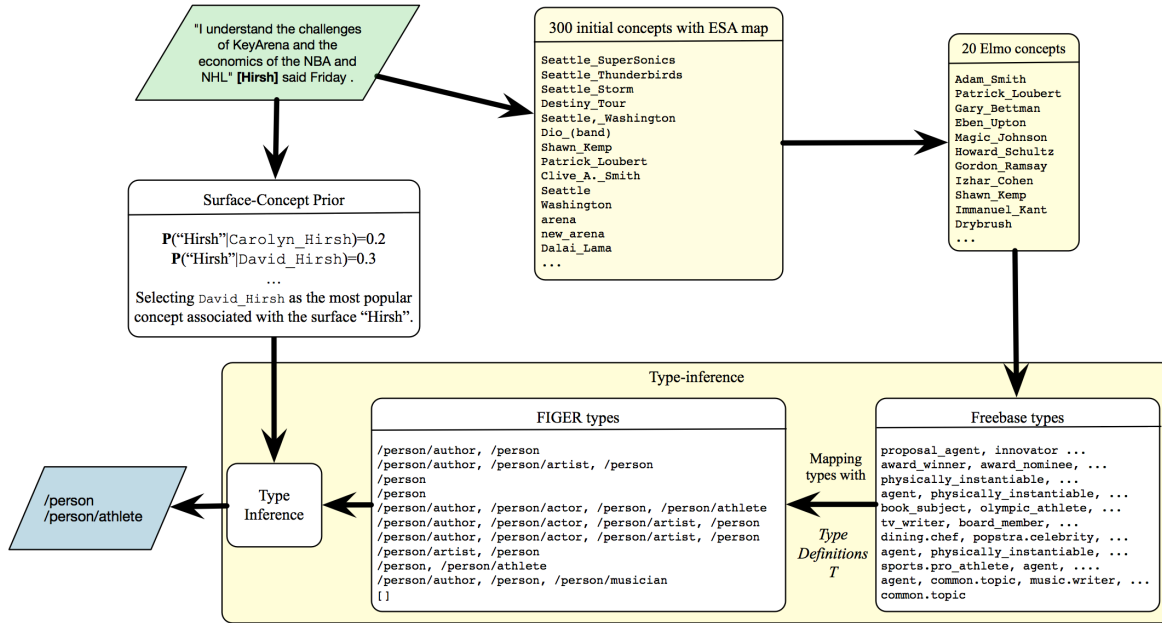


Figure 5: The output of each step in our system.

the type mapping function T . the following example, the mention gets mapped to an approximately correct concept `infantry`, but the system fails to map it to the correct type due to the limitations of the type definitions.

"When he left the [Army], Spencer got a job in Bozeman, where he used acupuncture to save a dog that couldn't walk anymore."

D Type Definitions

We include the type definitions used in the experiments here for completeness. The types on the left are for the target dataset, and the types on the right, are the types from FREEBASE, combined with logical operators AND (&&), OR (||), NOT (!), and any-type wildcard character (*).

For the type definitions of the FIGER dataset, we use the rules provided by (Ling and Weld, 2012), in addition to a few more definitions (Listing 1).

Listing 1: Additional definition used for FIGER beyond given mapping

```

/ORGANIZATION := /organization/organization
/LOCATION := /LOCATION || /BUILDING || /
transportation/road
/ORGANIZATION/COMPANY := (/ORGANIZATION/
COMPANY || /NEWS_AGENCY) && !(/
ORGANIZATION/EDUCATIONAL_INSTITUTION
|| /ORGANIZATION/SPORTS_LEAGUE)
/WRITTEN_WORK := /WRITTEN_WORK && !/
NEWS_AGENCY

```

```

/ART := /ART || /WRITTEN_WORK

```

Listing 2: Type definition used of the BBN set.

```

/ORGANIZATION := /ORGANIZATION/
ORGANIZATION
/PERSON := /PEOPLE/PERSON
/PLANT := /BASE/PLANTS/PLANT
/BUILDING := /ARCHITECTURE/BUILDING
/DISEASE := /MEDICINE/DISEASE
/LANGUAGE := /LANGUAGE/HUMAN_LANGUAGE
/LAW := /LAW && !/ORGANIZATION
/ANIMAL := /BIOLOGY/ANIMAL
/GPE/CITY := /LOCATION/CITYTOWN
/GPE/COUNTRY := /LOCATION/COUNTRY
/GPE/STATE_PROVINCE := /LOCATION/
CN_PROVINCE || /BASE/LOCATIONS/
STATES_AND_PROVENCES
/LOCATION := /LOCATION/LOCATION
/LOCATION/CONTINENT := /LOCATION/CONTINENT
|| /BASE/LOCATIONS/CONTINENTS
/LOCATION/RIVER := /GEOGRAPHY/RIVER
/LOCATION/LAKE_SEA_OCEAN := /GEOGRAPHY/
BODY_OF_WATER && !/LOCATION/RIVER
/LOCATION/REGION := /LOCATION/
STATISTICAL_REGION
/FAC/AIRPORT := /AVIATION/AIRPORT
/FAC/HIGHWAY_STREET := /TRANSPORTATION/
ROAD
/FAC/BRIDGE := /TRANSPORTATION/BRIDGE
/GAME := /CVG/COMPUTER_VIDEOGAME
/PRODUCT/VEHICLE := /AUTOMOTIVE/MODEL || /
AVIATION/AIRCRAFT_MODEL
/PRODUCT/WEAPON := /LAW/INVENTION
/WORK_OF_ART/BOOK := /BOOK/WRITTEN_WORK
/WORK_OF_ART/SONG := /MUSIC/COMPOSITION
/WORK_OF_ART/PAINTING := /VISUAL_ART/
ARTWORK
/WORK_OF_ART/PLAY := /THEATER/PLAY
/EVENT := /TIME/EVENT
/EVENT/WAR := /MILITARY/WAR

```

```

/EVENT/HURRICANE := /METEOROLOGY/
TROPICAL_CYCLONE
/SUBSTANCE/FOOD := /FOOD/FOOD
/SUBSTANCE/DRUG := /MEDICINE/DRUG
/SUBSTANCE/CHEMICAL := /CHEMISTRY/
CHEMICAL_COMPOUND
/ORGANIZATION/HOTEL := /TRAVEL/HOTEL
/ORGANIZATION/HOSPITAL := /MEDICINE/
HOSPITAL
/ORGANIZATION/CORPORATION := /BUSINESS/
EMPLOYER && !/ORGANIZATION/
GOVERNMENT
/ORGANIZATION/POLITICAL := /GOVERNMENT/
POLITICAL_PARTY
/ORGANIZATION/RELIGIOUS := /RELIGION/
RELIGION && !/LOCATION/CONTINENT
/ORGANIZATION/EDUCATIONAL := /EDUCATION/
ACADEMIC_INSTITUTION || /EDUCATION/
EDUCATIONAL_INSTITUTION
/ORGANIZATION/GOVERNMENT := /GOVERNMENT/
GOVERNMENT_AGENCY || /GOVERNMENT/
GOVERNMENTAL_BODY || /GOVERNMENT/
GOVERNMENT

```

Listing 3: Type definition used of the OntoNotes_{fine} set.

```

/PERSON := /PEOPLE/PERSON
/PERSON/ARTIST/AUTHOR := /BOOK/AUTHOR
/PERSON/ARTIST/ACTOR := /FILM/ACTOR
/PERSON/ARTIST/MUSIC := /MUSIC/ARTIST
/PERSON/ATHLETE := /SPORTS/PRO_ATHLETE
/PERSON/DOCTOR /MEDICINE/PHYSICIAN
/PERSON/POLITICAL_FIGURE := /GOVERNMENT/
POLITICIAN
/PERSON/LEGAL := /BASE/CRIME/
CRIMINAL_DEFENCE_ATTORNEY || /BASE/
CRIME/LAWYER_TYPE || /LAW/JUDGE
/PERSON/TITLE := /FICTIONAL_UNIVERSE/
FICTIONAL_JOB_TITLE || /BUSINESS/JOB_TITLE
|| /GOVERNMENT/
GOVERNMENT_OFFICE_OR_TITLE || /
GOVERNMENT/
GOVERNMENT_OFFICE_CATEGORY
/LOCATION/STRUCTURE/AIRPORT := /AVIATION/
AIRPORT
/LOCATION/STRUCTURE := /ARCHITECTURE/
BUILDING
/LOCATION/STRUCTURE/HOTEL := /TRAVEL/HOTEL
/LOCATION/STRUCTURE/SPORTS_FACILITY := /
SPORTS/SPORTS_FACILITY
/LOCATION/GEOGRAPHY/BODY_OF_WATER := /
GEOGRAPHY/BODY_OF_WATER
/LOCATION/GEOGRAPHY/MOUNTAIN := /
GEOGRAPHY/MOUNTAIN
/LOCATION/GEOGRAPHY := /GEOGRAPHY/*
/LOCATION/TRANSIT/BRIDGE := /TRANSPORTATION
/BRIDGE
/LOCATION/TRANSIT/RAILWAY := /RAIL/RAILWAY
/LOCATION/TRANSIT/ROAD := /TRANSPORTATION/
ROAD
/LOCATION/CITY := /LOCATION/CITYTOWN
/LOCATION/COUNTRY := /LOCATION/COUNTRY
/LOCATION/PARK := /AMUSEMENT_PARKS/PARK || /
BASE/USNATIONALPARKS/US_NATIONAL_PARK
/LOCATION := /LOCATION/LOCATION
/ORGANIZATION := /ORGANIZATION/
ORGANIZATION_TYPE || /ORGANIZATION/
ORGANIZATION

```

```

/ORGANIZATION/COMPANY/NEWS := /BASE/
NEWSEVENTS/
NEWS_REPORTING_ORGANISATION || /BOOK/
PUBLISHING_COMPANY
/ORGANIZATION/COMPANY/BROADCAST := /
BROADCAST/PRODUCER
/ORGANIZATION/COMPANY := /BUSINESS/
EMPLOYER
/ORGANIZATION/EDUCATION := /EDUCATION/
ACADEMIC_INSTITUTION || /EDUCATION/
EDUCATIONAL_INSTITUTION
/ORGANIZATION/GOVERNMENT := /GOVERNMENT/
GOVERNMENT_AGENCY || /GOVERNMENT/
GOVERNMENT || /GOVERNMENT/
GOVERNMENTAL_BODY
/ORGANIZATION/MILITARY := /MILITARY/
MILITARY_UNIT
/ORGANIZATION/POLITICAL_PARTY := /
GOVERNMENT/POLITICAL_PARTY
/ORGANIZATION/SPORTS_TEAM := /SPORTS/
SPORTS_TEAM
/ORGANIZATION/STOCK_EXCHANGE := /FINANCE/
STOCK_EXCHANGE
/OTHER/ART/BROADCAST := /TV/TV_PROGRAM
/OTHER/ART/FILM := /FILM/FILM
/OTHER/ART/MUSIC := /MUSIC/ALBUM || /MUSIC/
COMPOSITION
/OTHER/ART/STAGE := /THEATER/PLAY || /OPERA/
OPERA
/OTHER/ART/WRITING := /BOOK/WRITTEN_WORK ||
/BOOK/SHORT_STORY || /BOOK/POEM || /BOOK/
LITERARY_SERIES || /BOOK/PUBLICATION
/OTHER/EVENT := /TIME/EVENT
/OTHER/EVENT/HOLIDAY := /TIME/HOLIDAY
/OTHER/EVENT/VIOLENT_CONFLICT := /MILITARY/
MILITARY_CONFLICT
/OTHER/HEALTH/TREATMENT := /MEDICINE/
MEDICAL_TREATMENT
/OTHER/AWARD := /AWARD/AWARD
/OTHER/BODY_PART := /MEDICINE/
ANATOMICAL_STRUCTURE
/OTHER/CURRENCY := /FINANCE/CURRENCY
/OTHER/LIVING_THING/ANIMAL := /BIOLOGY/
ANIMAL
/OTHER/LIVING_THING := /BASE/PLANTS/PLANT
/OTHER/PRODUCT/WEAPON := /LAW/INVENTION
/OTHER/PRODUCT/VEHICLE := /AUTOMOTIVE/
MODEL || /AVIATION/AIRCRAFT_MODEL
/OTHER/PRODUCT/COMPUTER := /COMPUTER/*
/OTHER/PRODUCT/SOFTWARE := /COMPUTER/
SOFTWARE
/OTHER/FOOD := /FOOD/FOOD
/OTHER/RELIGION := /RELIGION/RELIGION
/OTHER/HERITAGE := /PEOPLE/ETHNICITY
/OTHER/LEGAL := /USER/SPROCKETONLINE/
ECONOMICS/LEGISLATION || /USER/
TSEgaran/LEGAL/ACT_OF_CONGRESS || /
USER/SKUD/LEGAL/TREATY || /LAW/
CONSTITUTIONAL_AMENDMENT
/OTHER := !ALL_TYPES_EXCLUDING_OTHER* || /
OTHER*

```

Listing 4: Type definition used of the OntoNotes set.

```

/PERSON := /PEOPLE/* && /MUSIC/ARTIST
/LOC := /LOCATION/LOCATION
/ORG := /ORGANIZATION/* || /GOVERNMENT/
GOVERNMENT_BODY || /BUSINESS/EMPLOYER

```

```
|| /BOOK/NEWSPAPER || /RELIGION/RELIGION ||  
/MILITARY/MILITARY_COMBATANT
```

Listing 5: Type definition used of the MUC set.

```
/PER := /PEOPLE/* && /MUSIC/ARTIST  
/LOC := /LOCATION/LOCATION  
/ORG := /ORGANIZATION/* || /GOVERNMENT/  
GOVERNMENT_BODY || /BOOK/NEWSPAPER || /  
RELIGION/RELIGION
```

Listing 6: Type definition used of the CoNLL set.

```
/PER := /PEOPLE/PERSON  
/LOC := /LOCATION/LOCATION  
/ORG := /ORGANIZATION/ORGANIZATION
```

Listing 7: Type definition used of the BB3 set.

```
/BACTERIA := /*/MICROORGANISM/*
```
