

# Evaluating Class Membership Relations in Knowledge Graphs using Large Language Models

Bradley Allen and Paul Groth

ESWC 2024 Special Track on LLMs for KE

2024-05-29

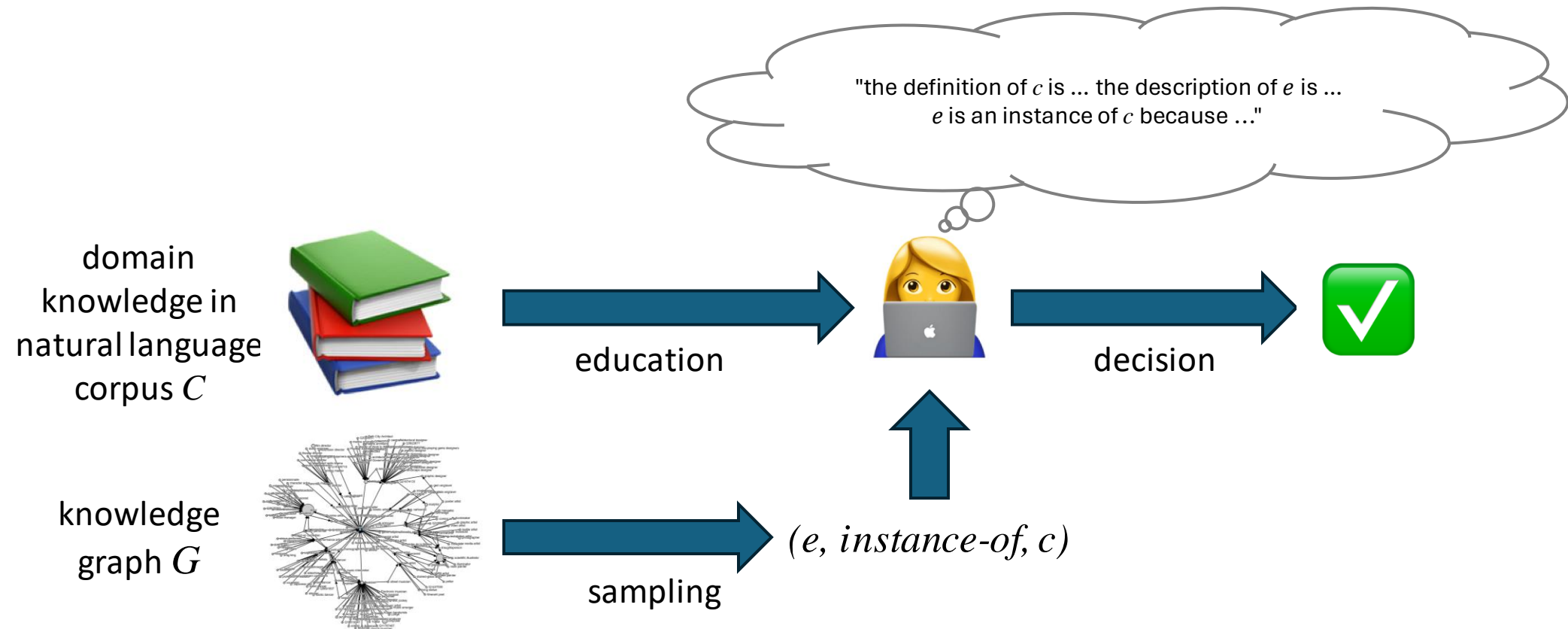
# Background

- Knowledge graph (KG) refinement is an increasingly important topic
- KG refinement is difficult
  - KGs can be messy (e.g., Wikidata quality issues)
  - Human evaluation is expensive
- Machine learning (ML) and natural language processing (NLP) may help by enabling automation of KG evaluation and refinement tasks
- Large language models (LLMs) are state-of-the-art ML/NLP technology with a broad range of application
- Can LLMs aid in KG refinement?

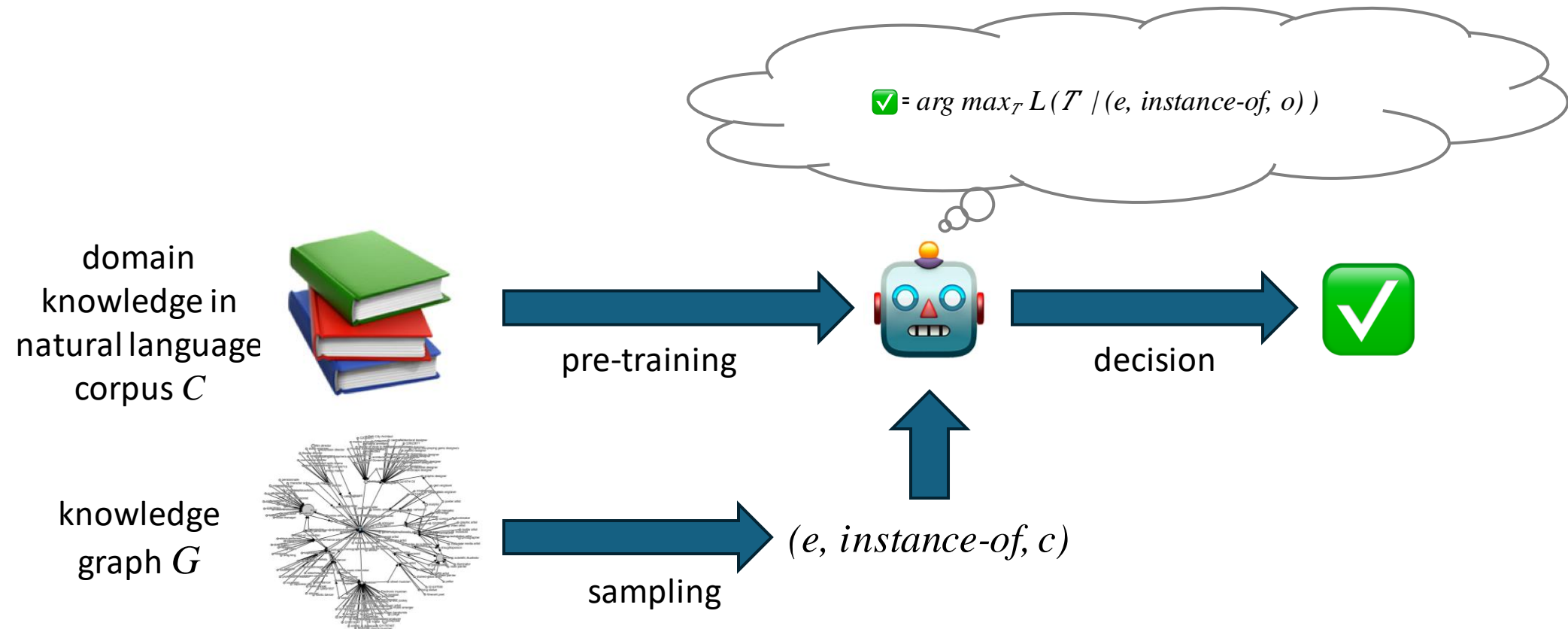
# Problem statement

- We will focus on how LLMs can be used to support the evaluation of *class membership relations* in a KG
  - Class membership relations are a principal way in which KGs represent classification schemes
  - Classification schemes are a crucial aspect of knowledge infrastructure, often with significant implications for social policy and scientific consensus
  - Class membership is also a key concern in data governance, which depends on "providing a set of mappings from a representation language to agreed-upon concepts in the real world" [Khatri and Brown]

# Class membership relation evaluation by a human reviewer



# Class membership relation evaluation by an LLM



# Approach

- Define a zero-shot chain-of-thought (CoT) classifier that takes a class and an entity from a KG and generates a rationale for whether the entity is a member of the class
- Apply the classifier to class/entity pairs from a KG and, treating the KG as ground truth, compute the resulting confusion matrices
- Perform a human review of classification errors to determine if the LLM or the KG is at fault

# Related work

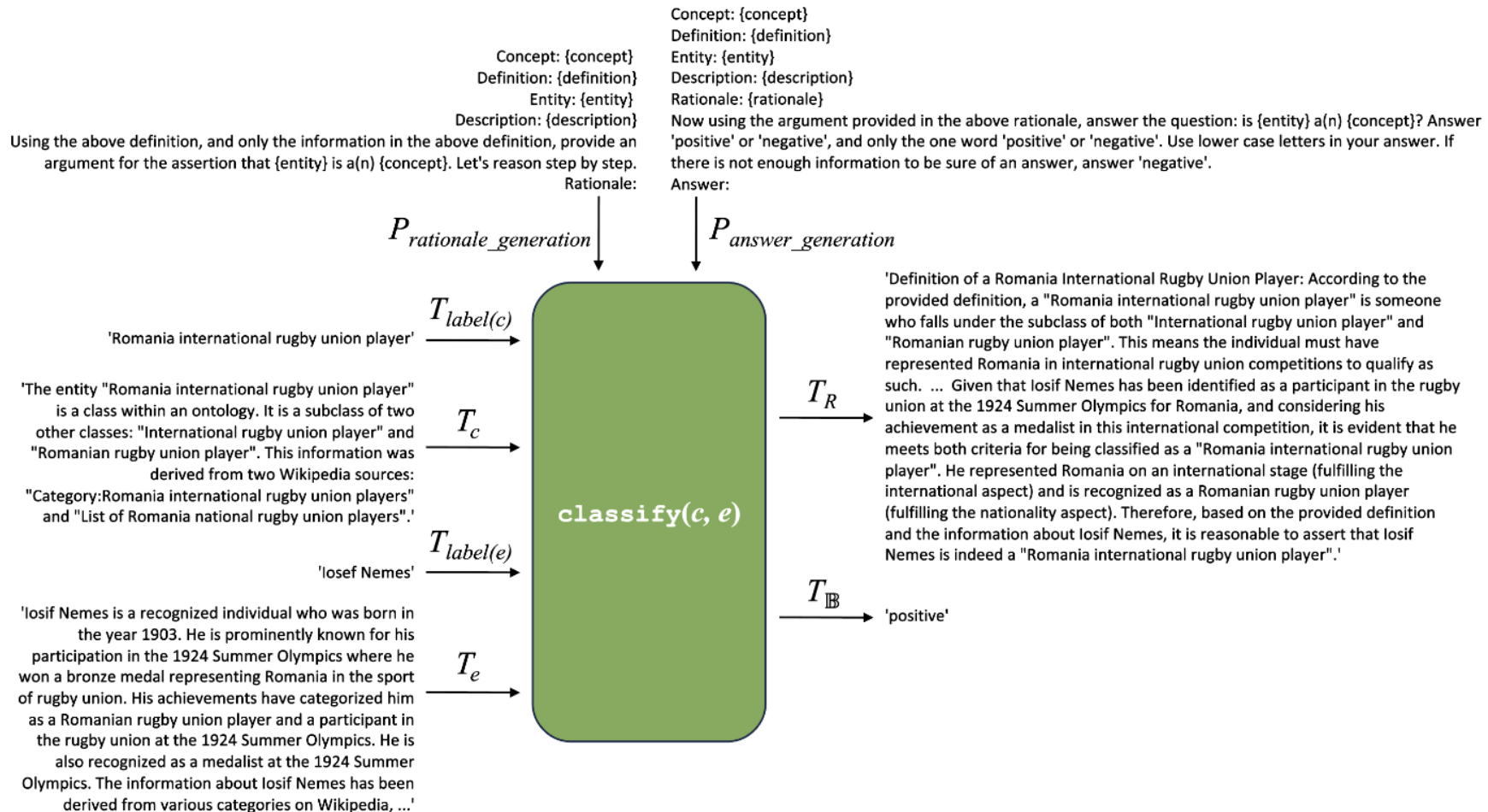
- LLMs for KE [Allen et al]
- KG refinement [Paulheim], [Xue & Zou], [Hofer et al]
- Automated fact checking [Guo et al], [Atanasova et al]

# Definitions

$e$	An entity in a knowledge graph $G$	$T_R$	A sequence of tokens providing a natural language rationale for a classification of $e$ given $c$
$c$	A class in a knowledge graph $G$	$T_{\mathbb{B}}$	A token providing a natural language classification of $e$ given $c$
$T_c$	A sequence of tokens that describe $c$	$\mathbf{ext}(c)$	The extension of $c$
$T_e$	A sequence of tokens that describe $e$	$\mathbf{classify}(c, e)$	A function returning $T_{\mathbb{B}}, T_R$ given $c, e$
$T_{label(e)}$	A sequence of tokens that is the <b><i>rdfs:label</i></b> of $e$	$P_{rationale\_generation}$	A prompt to generate $T_R$
		$P_{answer\_generation}$	A prompt to generate $T_{\mathbb{B}}$



# A zero-shot chain-of-thought classifier that generates rationales for its decisions



# Research questions

- **Q1:** Can the classifiers exhibit good alignment with KGs?
- **Q2:** Can the classifiers detect missing or incorrect relations?

# Experiments

**input** : a pair of classes  $c, d$  from  $G \mid (c, \text{subClassOf}, d) \in G$

**output**: a confusion matrix  $M$

$(TP, FP, TN, FN) \leftarrow (0, 0, 0, 0);$

$E^+ \leftarrow$  a sample from  $\text{ext}(c);$

$E^- \leftarrow$  a sample from  $\text{ext}(d) \setminus \text{ext}(c);$

**foreach**  $e \in E^+$  **do**

$(T_R, T_B) \leftarrow \text{classify}(c, e);$

**if**  $T_B = \textit{positive}$  **then**  $TP \leftarrow TP + 1;$

**else**  $FP \leftarrow FP + 1;$

**end**

**foreach**  $e \in E^-$  **do**

$(T_R, T_B) \leftarrow \text{classify}(c, e);$

**if**  $T_B = \textit{negative}$  **then**  $TN \leftarrow TN + 1;$

**else**  $FN \leftarrow FN + 1;$

**end**

$M \leftarrow [[TP, FP], [FN, TN]];$

- LLMs:
  - gpt-4-0125-preview
  - gpt-3.5-turbo
  - gemma-7b-it
  - gemma-2b-it
  - Mixtral-8x7B-Instruct-v0.1
  - Mistral-7B-Instruct-v0.2
  - Llama-2-70b-chat-hf
- KGs: Wikidata, CaLiGraph
- Data:
  - 20 Wikidata classes, 19 CaLiGraph classes randomly sampled
  - 20 positive and ~20 negative examples randomly sampled per class
  - Serializations limited to 20 triples per example
  - ~800 class/entity pairs per KG
- Error analysis: manually review FNs, FPs with rationales and assign error to LLM or KG
  - LLM errors: incorrect reasoning, missing data
  - KG errors: missing relation, incorrect relation
  - Error analysis performed for gpt-4-0125-preview

# Performance metrics

KG	LLM	ACC	AUC	F1	$\kappa$
Wikidata	gpt-4-0125-preview	<b>0.830</b>	<b>0.830</b>	<b>0.823</b>	<b>0.660</b>
	gemma-7b-it	0.726	0.727	0.705	0.454
	Mixtral-8x7B-Instruct-v0.1	0.697	0.696	0.654	0.393
	Mistral-7B-Instruct-v0.2	0.671	0.671	0.620	0.342
	gemma-2b-it	0.674	0.670	0.629	0.330
	gpt-3.5-turbo	0.627	0.627	0.547	0.255
	Llama-2-70b-chat-hf	0.631	0.616	0.569	0.239
CaLiGraph	gpt-4-0125-preview	<b>0.900</b>	<b>0.893</b>	<b>0.889</b>	<b>0.788</b>
	Mixtral-8x7B-Instruct-v0.1	0.893	0.884	0.874	0.767
	gpt-3.5-turbo	0.842	0.833	0.815	0.665
	Mistral-7B-Instruct-v0.2	0.812	0.803	0.779	0.605
	gemma-7b-it	0.783	0.774	0.750	0.547
	Llama-2-70b-chat-hf	0.637	0.625	0.558	0.252
	gemma-2b-it	0.563	0.543	0.422	0.090

- Classifiers can exhibit good alignment with KGs (**Q1**)
  - One LLM was in moderate agreement ( $\kappa > 0.60$ ) with Wikidata
  - Four were in moderate agreement with CaLiGraph

# Error analysis of gpt-4-0125-preview results

KG	$N$	$FP$	$FN$	human-KG $\kappa$	human-LLM $\kappa$	missing data	missing relation	incorrect relation	incorrect reasoning
Wikidata	136	46	90	<b>0.243</b>	-0.241	34 (25.0%)	15 (11.0%)	33 (24.3%)	54 (39.7%)
CaLiGraph	77	27	50	-0.295	<b>0.198</b>	28 (36.4%)	19 (24.7%)	20 (26.0%)	10 (13.0%)
	213	73	140			62 (29.1%)	34 (16.0%)	53 (24.9%)	64 (30.0%)

- Error analysis based on review by one of the authors
- Classifiers can detect missing or incorrect relations (**Q2**)
  - 40.9% of errors were due to the problems with the KG
  - 29.1% of errors were due to missing or insufficient data in the entity description
  - 30.0% of errors due to incorrect reasoning by the LLM
- Pairwise human-KG and human-LLM agreement differed between the KGs
  - Human showed fair agreement with Wikidata and no agreement with the classifier
  - Human showed slight agreement with the classifier and no agreement with CaLiGraph

# Incorrect relation (Wikidata)

$T_c$

**Village (Japan):** A village (村, mura) is a local administrative unit in Japan. It is a local public body along with prefecture (県, ken, or other equivalents), city (市, shi), and town (町, chō, sometimes machi). Geographically, a village's extent is contained within a prefecture. It is larger than an actual settlement, being in actuality a subdivision of a rural district (郡, gun), which are subdivided into towns and villages with no overlap and no uncovered area. As a result of mergers and elevation to higher statuses, the number of villages in Japan is decreasing. As of 2006, 13 prefectures no longer have any villages: Tochigi (since March 20, 2006), Fukui (since March 3, 2006), Ishikawa (since March 1, 2005), Shizuoka (since July 1, 2005), Hyōgo (since April 1, 1999), Mie (since November 1, 2005), Shiga (since January 1, 2005), Hiroshima (since November 5, 2004), Yamaguchi (since March 20, 2006), Ehime (since January 16, 2005), Kagawa (since April 1, 1999), Nagasaki (since October 1, 2005), and Saga (since March 20, 2006). Atarashiki-mura (which is an autonomous village community) is not included in the list below.

$T_e$

**Tachikawa:** Tachikawa (立川市, Tachikawa-shi) is a city located in the western portion of Tokyo Metropolis, Japan. As of 1 March 2021, the city had an estimated population of 184,383 in 93,428 households, and a population density of 7600 persons per km<sup>2</sup>. The total area of the city is 24.36 square kilometres (9.41 sq mi).

$T_{\mathbb{B}}$

negative

$T_R$

Based on the information provided and the definition of a village in Japan, it is not possible to argue that Tachikawa is a village. Here's why, step by step:

- Definition of a Village in Japan:** A village in Japan, as defined, is a local administrative unit that is a subdivision of a rural district (郡, gun). It is one of the types of local public bodies, alongside prefecture, city, and town. Villages are characterized by being part of the rural district's administrative structure.
- Status of Tachikawa:** Tachikawa is explicitly described as a city (立川市, Tachikawa-shi). In Japan, the designation of "city" is distinct from that of a "village". Cities and villages are different types of administrative units with specific criteria for their classification, including population size, administrative structure, and economic functions.
- Population and Density:** The population and density of Tachikawa, with an estimated population of 184,383 and a population density of 7600 persons per km<sup>2</sup>, far exceed what would typically be expected of a village. Villages, by their nature, tend to have smaller populations and lower population densities, reflecting their rural character.
- Geographical and Administrative Context:** Tachikawa's location within the Tokyo Metropolis and its designation as a city indicate that it is part of a highly urbanized and densely populated area. This context is inconsistent with the typical geographical and administrative setting of a village, which would be within a rural district (郡, gun).
- Lack of Overlap with Rural Districts:** The definition of a village mentions that it is a subdivision of a rural district with no overlap and no uncovered area. Tachikawa, being a city, does not fit this criterion as it is not a subdivision of a rural district but rather an independent administrative unit within the Tokyo Metropolis.

In conclusion, based on the provided definition of a village in Japan and the specific details about Tachikawa, it is clear that **Tachikawa cannot be classified as a village**. Its status as a city, along with its population size, density, and urban context, directly contradicts the criteria for being considered a village in Japan.

# Missing relation (CaLiGraph)

$T_c$

The entity "**State leader in 2014**" is a class within the Caligraph knowledge graph's ontology. It is a specific subclass of the broader "State leader" class. The information for this entity is derived from the Wikipedia page titled "List of state leaders in 2014".

This class includes several individuals who were state leaders in 2014. These individuals are 'Akilisi Pōhiva, Armando Guebuza, Barry Haase, Hovik Abrahamyan, Michael Dunkley, Miloš Zeman, Nicos Anastasiades, Omar al-Hassi, Tsakhiagiin Elbegdorj, Adam Wood, Beji Caid Essebsi, Harold Martin, Jorge Carlos Fonseca, Joyce Banda, Mikhail Myasnikovich, and Hassan Rouhani. Each of these individuals is classified as an instance of the "State leader in 2014" class, indicating their role as state leaders during that year.

$T_e$

**Zlatko Lagumdžija**, also known as "Lagumdžija", is a Bosniak individual who was born in 1955 and is currently living. He hails from Sarajevo, Bosnia and Herzegovina. He has had a significant political career, serving as the Foreign Minister of Bosnia and Herzegovina in several years, including 2001, 2002, 2012, 2013, 2014, and 2015. In addition to his role as Foreign Minister, he also held a position as a state leader in 2001 and 2002. His contributions and roles have led to his recognition as a notable Bosniak of Bosnia and Herzegovina and a prominent politician from Sarajevo.

$T_{\mathbb{B}}$

positive

$T_R$

**1. Definition of "State leader in 2014":** The class "State leader in 2014" within the Caligraph knowledge graph's ontology is defined as a subclass of "State leader" that specifically includes individuals who were state leaders in the year 2014. This class is populated based on the roles these individuals held during that year, as derived from the Wikipedia page titled "List of state leaders in 2014".

**2. Role of Zlatko Lagumdžija in 2014:** According to the provided information, Zlatko Lagumdžija served as the Foreign Minister of Bosnia and Herzegovina in several years, including 2014. The role of a Foreign Minister is a significant position within the government, involving the oversight of international relations and diplomacy on behalf of the state. This position is typically considered a leadership role within the state's executive branch or government.

**3. Criteria for being a "State leader":** The role of Foreign Minister inherently involves leadership responsibilities at the state level, particularly in the realm of international affairs. Given that the class "State leader" would encompass individuals who held significant leadership positions within their respective states, serving as a Foreign Minister qualifies an individual as a state leader.

**4. Conclusion:** Based on the definition of "State leader in 2014" as individuals who were state leaders during that year, and given that Zlatko Lagumdžija served as the Foreign Minister of Bosnia and Herzegovina in 2014, it logically follows that Zlatko Lagumdžija meets the criteria to be classified as a "State leader in 2014". His role as Foreign Minister in 2014 directly aligns with the criteria for inclusion in this specific class within the Caligraph knowledge graph's ontology. **Therefore, Zlatko Lagumdžija is indeed a "State leader in 2014" according to the provided definition and information.**

# Missing data (CaLiGraph)

$T_c$

The entity "**FK Sarajevo player**" is a class within the ontology, specifically labeled as "FK Sarajevo player". It is a subclass of several other classes, including "RestrictionHasValue\_team\_FK Sarajevo", "Footballer in Bosnia and Herzegovina", and "Association football player in Bosnia and Herzegovina". This suggests that the class "FK Sarajevo player" is a specific type of football player who is associated with the team FK Sarajevo in Bosnia and Herzegovina.

The information for this entity was derived from two Wikipedia pages: "Category:FK Sarajevo players" and "List of FK Sarajevo players". This suggests that the data for this class was gathered from these pages and includes all players listed there.

Several individuals are classified under this entity, indicating that they are or were players for FK Sarajevo. These individuals include Branislav Arsenijević, Sead Jesenković, Darko Raca, Franjo Vuleta, Ibrahim Biogradlić, Vule Trivunović, Sretko Vuksanović, Vukašin Višnjevac, Marko Savić, Frank Liivak, Matej Kochan, Safet Nadarević, and Said Fazlagić. Each of these individuals is classified as an "FK Sarajevo player", reinforcing their association with this specific football team.

$T_e$

**Darko Raca** is a named individual, born in 1977, who is a professional football defender from Bosnia and Herzegovina. He has played for several football clubs including FK Sloboda Tuzla, FK ČSK Čelarevo, FK Kozara Gradiška, and FK Laktaši. His career has seen him play in various leagues such as the Serbian First League, Division 2 of Swedish football, and Ettan Fotboll. Darko Raca is not only recognized as an expatriate footballer who has played in Serbia, Serbia and Montenegro, and Sweden, but also as a Bosnia and Herzegovina expatriate sports person in these countries. He is further categorized as a foreign football player in Serbia. Darko Raca, a living person, hails from Sarajevo.

$T_{\mathbb{B}}$

negative

$T_R$

1. **Identification of the Entity Class:** The entity class "FK Sarajevo player" is defined as a specific type of football player who is associated with the team FK Sarajevo in Bosnia and Herzegovina. This class is a subclass of broader categories such as "Footballer in Bosnia and Herzegovina" and "Association football player in Bosnia and Herzegovina", but it specifies the team association to FK Sarajevo.

2. **Inclusion Criteria for the Entity Class:** To be classified under the "FK Sarajevo player" entity, an individual must have an association with the FK Sarajevo football team. This association is typically through playing for the team, as indicated by the sources of information (Wikipedia pages on FK Sarajevo players).

3. **Analysis of Darko Raca's Profile:** Darko Raca is identified as a professional football defender from Bosnia and Herzegovina, with a career that includes playing for several football clubs. The clubs mentioned are FK Sloboda Tuzla, FK ČSK Čelarevo, FK Kozara Gradiška, and FK Laktaši. His career spans various leagues, including the Serbian First League and Division 2 of Swedish football.

4. **Lack of Direct Mention of FK Sarajevo:** In the provided information about Darko Raca, there is no direct mention of him playing for FK Sarajevo. His association with other clubs is detailed, but FK Sarajevo is not listed among the clubs he has played for.

5. **Conclusion Based on Provided Information:** Based solely on the information provided in the definition of "FK Sarajevo player" and the detailed profile of Darko Raca, there is no evidence to directly assert that Darko Raca is an FK Sarajevo player. The assertion that an individual is an FK Sarajevo player requires evidence of their association with FK Sarajevo, specifically playing for the team, which is not present in the information given about Darko Raca.

6. **Final Argument:** Therefore, without additional information linking Darko Raca to FK Sarajevo directly, such as records of him playing for the team or being listed on the team's roster, **we cannot logically conclude that Darko Raca fits the classification of an "FK Sarajevo player"** based on the provided definitions and his profile information.



# Incorrect reasoning (Wikidata)

 $T_c$ 

**History of cities:** Towns and cities have a long history, although opinions vary on which ancient settlements are truly cities. The benefits of dense settlement included reduced transport costs, exchange of ideas, sharing of natural resources, large local markets, and in some cases amenities such as running water and sewerage. Possible costs would include higher rate of crime, higher mortality rates, higher cost of living, worse pollution, traffic and high commuting times. Cities grow when the benefits of proximity between people and firms are higher than the cost.

 $T_e$ 

**military history of Japan:** The military history of Japan covers a vast time-period of over three millennia - from the Jōmon (c. 1000 BC) to the present day. After a long period of clan warfare until the 12th century, there followed feudal wars that culminated in military governments known as the Shogunate. Japanese history records that a military class and the Shōgun ruled Japan for 676 years - from 1192 until 1868. The Shōgun and the samurai warriors stood near the apex of the Japanese social structure - only the aristocratic nobility nominally outranked them. The sakoku policy effectively closed Japan from foreign influences for 212 years - from 1641 to 1853. Feudal militarism transitioned to imperialism in the 19th century after the arrival of Admiral Perry in 1853 and the elevation of Emperor Meiji in 1868. Western colonial powers and their imperialist policies impacted on Japan's outlook and led to Japanese colonialism and rampant imperialism (c. 1895 - 1945) until Japan's defeat in World War II. The 1947 Japanese Constitution prohibits Japan from offensively using war against other nations. This led to the establishment of the Japan Self-Defense Forces in 1954. The U.S.–Japan Alliance (1951 onwards) requires the United States of America to protect Japan and to conduct offensive duties. In 2015 the Constitution was reinterpreted to allow collective self-defense of Japan's allies. As of 1954 the Japan Self-Defense Forces (JSDF) consist of the Japan Ground Self-Defense Force (JGSDF), Japan Maritime Self-Defense Force (JMSDF) and Japan Air Self-Defense Force (JASDF). The Prime Minister is the commander-in-chief of the Japan Self-Defense Forces. Military authority runs from the Prime Minister to the cabinet-level Minister of Defense of the Japanese Ministry of Defense. The Prime Minister and Minister of Defense are advised by the Chief of Staff, Joint Staff, who heads the Joint Staff (統合幕僚監部, Tōgō Bakuryō Kanbu). The Chief of Staff, Joint Staff, is the highest-ranking military officer in the Japan Self-Defense Forces, and is the head of the Operational Authority over the JSDF, executing orders of the Minister of Defense with directions from the Prime Minister.

 $T_{\mathbb{B}}$ 

**positive**

To argue that the military history of Japan is a history of cities, we need to draw connections between the development, significance, and transformation of Japanese cities and their military history. The information provided offers several avenues to establish this relationship:

- 1. Feudal Wars and the Shogunate:** The period of clan warfare until the 12th century and the subsequent feudal wars that led to the establishment of military governments known as the Shogunate had profound impacts on the development and structure of Japanese cities. Cities during these times often centered around castles or fortifications, serving as administrative and military hubs. The Shogunate system, with its military class and samurai warriors, influenced the social and physical structure of cities, prioritizing defense and the needs of the ruling military elite.
- 2. Sakoku Policy:** The sakoku policy, which effectively closed Japan from foreign influences for 212 years, had significant implications for Japanese cities. This isolationist policy led to a period of internal peace and stability, allowing for urban development and growth without the external military threats that characterized earlier periods. Cities became centers of culture, commerce, and internal governance, reflecting the shift in focus from military defense to internal administration and development.
- 3. Transition to Imperialism:** The arrival of Admiral Perry and the elevation of Emperor Meiji marked the end of feudal militarism and the beginning of imperialism. This transition had a transformative effect on Japanese cities, as the country rapidly industrialized and modernized. Cities expanded and modernized, reflecting the new imperial ambitions and the influence of Western colonial powers. The growth of cities during this period was directly tied to Japan's military and imperial ambitions, as urban centers became crucial for industrial production, military recruitment, and administration of the growing empire.
- 4. Post-World War II Reconstruction and the Japan Self-Defense Forces:** Following Japan's defeat in World War II, the country underwent a period of reconstruction that again reshaped its cities. The establishment of the Japan Self-Defense Forces and the reinterpretation of the Constitution to allow collective self-defense led to the development of military bases and facilities within and around cities. The presence of the JSDF and the U.S.–Japan Alliance influenced the strategic importance of certain urban areas and their development post-war.
- 5. Modern Urban Development and Military Influence:** The structure and development of modern Japanese cities cannot be fully understood without considering their military history. From the feudal castles that formed the nuclei of historical cities to the modern bases of the JSDF that impact local economies and urban landscapes, the military history of Japan is deeply intertwined with the history of its cities. In conclusion, **the military history of Japan is indeed a history of cities**, as it has significantly influenced the development, structure, and transformation of Japanese urban centers throughout the centuries. From the feudal periods to modern times, military needs, policies, and transformations have played a central role in shaping Japanese cities, making the military history of Japan an integral part of the broader history of its urban development.

 $T_R$

# Discussion

- What is a key challenge going forward that your work gives rise to?
  - *Metalinguistic disagreement* between an LLM and a KG presents a problem for the use of LLMs for text-to-fact extraction and fact-to-text verbalization
  - LLMs for KE must address *metalinguistic negotiation* to resolve this challenge
  - Our work shows how to
    - Identify concepts where metalinguistic disagreement exists
    - Quantify the degree of disagreement in terms of inter-annotator alignment
- What did you wish you knew before starting this work?
  - The paucity of usable natural language descriptions of knowledge graph entities and classes (surprisingly, even in the context of Wikidata/Wikipedia)

# Limitations

- Robustness, response time of KG SPARQL endpoints made data sampling challenging
- Limit of 20 triples/serialization, lack of detail in Wikidata entity descriptions contributed to LLM errors
- Cost, response time, token/rate limiting of LLM API led to limited number of experiments

# Future work

- Extend approach to support evaluation of  $k$ -ary relations using a first-order intensional semantics based on knowledge probing
- Increase the number of human annotators in support of error analysis and investigation of metalinguistic disagreement between KGs and LLMs



UNIVERSITEIT  
VAN AMSTERDAM

INDE lab

**Thank you!**

Code and data available on Github

<https://github.com/bradleypallen/evaluating-kg-class-memberships-using-llms>