FACULTY OF APPLIED SCIENCES
UNIVERSITY
OF WEST BOHEMIA

3D Land Administration

12th International FIG Workshop on LADM & 3D LA
24-26 September 2024, Kuching, Malaysia

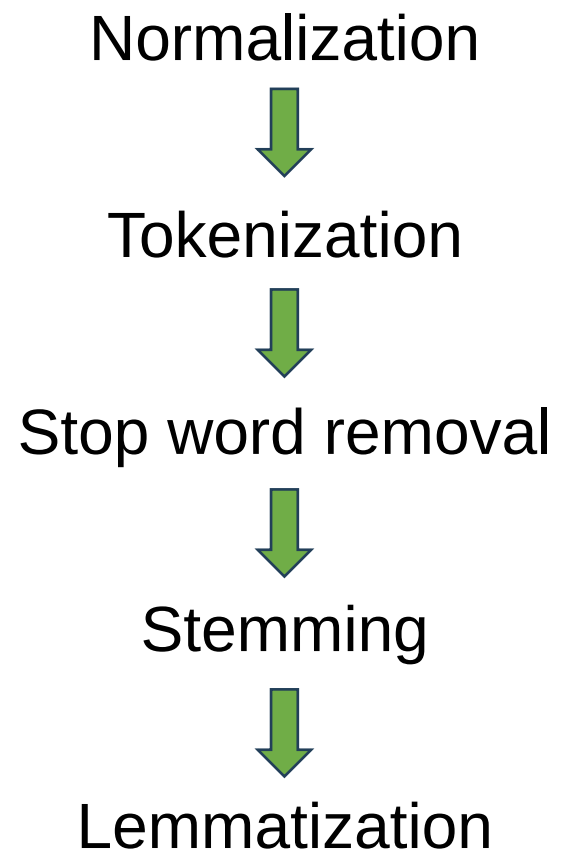
Determining the semantic similarity of definitions by artificial intelligence for the needs of 3D Land Administration: a case of building units

Martin VANĚK, Karel JANEČKA, and Otakar ČERBA
Czech Republic, University of West Bohemia, Faculty of Applied Sciences,
Department of Geomatics

Motivation

- different definitions for individual terms from the field of 3D Land Administration
- prevent future misunderstandings
- conversion between standards (e.g. IFC and CityGML)
- AI undergone significant development
- AI includes a model for calculating semantic similarity

Natural Language Processing (NLP)



Vectorization

Today is a beautiful day ⇨ [-0.005, 0.012, …, -0.025]

- example of models used to vectorize text:

Bag of Words (BoG)

Word2vec

Global Vector for word representation (GloVe)

Computing semantic similarity

- Euclidean distance

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

- Cosine similarity

$$\cos\theta = \frac{\mathbf{p}\cdot\mathbf{q}}{\|\mathbf{p}\|\|\mathbf{q}\|} = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2}\sqrt{\sum_{i=1}^{n} q_i^2}}$$

Selection of terms and definitions

- Land Administration Domain Model (LADM) was chosen as the basic standard, the terms of which will be compared with the terms of other standards
- standards for comparison:

Industry Foundation Classes (IFC 4.3)

Land and Infrastructure Conceptual Model Standard (LandInfra)

City Geography Markup Language (CityGML 3.0)

Selection of terms and definitions

Land Administration Domain Model (LADM)		
1	basic administrative unit	Administrative entity, subject to registration (by law), or recordation [by informal right, or customary right, or another social tenure relationship], consisting of zero or more spatial units against which (one or more) unique and homogeneous rights [e.g. ownership right or land use right], responsibilities or restrictions are associated to the whole entity, as included in a land administration system.
2	boundary	Set that represents the limit of an entity.
3	boundary face	Face that is used in the 3-dimensional representation of a boundary of a spatial unit.
4	boundary face string	Boundary forming part of the outside of a spatial unit.
5	building unit	Component of building (the legal, recorded or informal space of the physical entity).
6	land	The surface of the Earth, the materials beneath, the air above and all things fixed to the soil.
7	spatial unit	Single area (or multiple areas) of land and/or water, or a single volume (or multiple volumes) of space.

Semantic similarity results

OpenAI

Land Administration Domain Model (LADM)

IFC 4.3

		1	2	3	4	5	6	7
		basic administrative unit	boundary	boundary face	boundary face string	building unit	land	spatial unit
1	building information modelling	76,98	70,95	76,28	74,42	79,34	72,99	77,41
2	element	77,91	74,00	81,03	80,45	81,16	79,88	80,84
3	entity	77,92	73,32	73,99	74,63	78,18	72,84	76,09
4	facility	79,86	70,83	79,30	81,43	85,29	78,98	83,41
5	feature	73,76	70,19	76,83	76,50	76,62	71,97	74,41
6	model	74,57	75,70	71,09	71,81	72,88	69,22	72,61
7	object	77,89	72,57	79,43	81,62	82,86	81,65	81,79
8	product	75,45	69,25	74,96	75,07	77,44	78,79	78,05
9	property	76,64	73,57	77,38	76,24	75,50	73,27	76,86
10	property set	78,78	75,36	76,78	77,57	77,45	76,13	79,85
11	representation	80,71	72,27	75,66	76,96	79,06	74,12	81,15
12	space	80,21	78,40	84,51	83,64	80,84	79,49	85,41

Semantic similarity results

Hugging Face

Land Administration Domain Model (LADM)

IFC 4.3

		1	2	3	4	5	6	7
		basic administrative unit	boundary	boundary face	boundary face string	building unit	land	spatial unit
1	building information modelling	22,40	12,30	19,40	13,80	43,10	1,80	18,30
2	element	24,60	23,90	39,90	33,00	41,30	24,90	22,80
3	entity	29,60	38,20	12,80	10,40	31,30	12,10	22,20
4	facility	22,60	19,70	19,80	27,20	57,50	36,40	38,50
5	feature	9,00	4,80	32,30	19,70	34,00	14,40	13,30
6	model	39,60	47,00	7,10	6,60	26,90	6,20	17,60
7	object	11,40	22,20	18,30	25,10	21,40	22,60	32,20
8	product	16,40	13,30	9,00	17,40	35,70	25,30	7,50
9	property	30,40	31,10	35,80	21,70	34,70	12,30	16,20
10	property set	29,20	45,90	23,60	23,80	40,10	19,30	27,90
11	representation	27,40	26,10	13,40	8,30	35,50	14,10	25,90
12	space	21,00	27,70	54,40	35,50	31,90	15,10	39,20

Comparison of results

- both AIs determined the same two definitions in 71.4% of cases as the definitions with the greatest similarity
- both AIs determined the same two definitions in 19% of cases as the definitions with the lowest similarity
- if the semantic similarity is low for one AI, it can be expected to be low for the other, and vice versa
- Even if different standards use the same term, it does not necessarily mean that the definitions of the term will have a high semantic similarity
- BoG method for calculating semantic similarity is highly inaccurate

Conclusion

- calculation of the semantic similarity of the texts was successful
- can be expected that if one AI calculate a low similarity value between two definitions, then the other AI will also calculate a low similarity value for these definitions, and vice versa
- more appropriate to use calculations that also consider the context
- artificial intelligence can help with comparing selected terms connected to the building units across different standards
- it may happen that the same terms are not the closest in meaning in the two standards
- furthermore, we would like to carry out research that would verify the usability of the calculation of semantic similarity by AI for the needs of class mapping between individual standards

Acknowledgements

- the first author of the paper was supported by the project SGS-2022-027, Use of Mathematics and Informatics in Geomatics V

FACULTY OF APPLIED SCIENCES
UNIVERSITY
OF WEST BOHEMIA

Thank you for your attention

Martin Vaněk / vanekma@kgm.zcu.cz