

Name	Naoki Yoshinaga	Laboratory	Institute of Industrial Science (IIS)	Research	Natural Language Processing
-------------	-----------------	-------------------	---------------------------------------	-----------------	-----------------------------

Efficient and Accurate Natural Language Processing

Languages enable us to think, record experiences, and communicate. Our laboratory explores efficient and accurate **natural language processing (NLP)** to understand the human mind and society through countless words and support various language activities. The pursuit of fast, compact, and accurate models reveals **the “shape” of language, as a natural phenomenon, that behaves both probabilistically and regularly (computational linguistics)**, ultimately leading to the understanding and refinement of human intelligence.

<https://www.tkl.iis.u-tokyo.ac.jp/~ynaga/index.en.html>

We think studies by students with different linguistic, cultural, and academic backgrounds, from their perspectives, elucidate languages; **students are able and expected to voluntarily set their research themes**. We welcome those who tackle novel tasks, not tied to accuracy-driven studies on worn-out benchmarks.

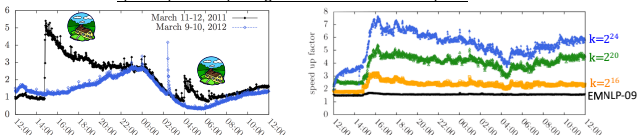
Read *all* language in the wild

Humans generate vast amounts of real-world information in microblogs (**social big data**), where both quality and quantity change over time. We leverage data structures and algorithms from computer science to achieve **the world’s fastest NLP**. Specifically, we have developed a classifier that speeds up as text redundancy increases and implemented a core text processing capable of handling 1,000,000 sentences per second. We have also established a method to instantly detect the emergence and disappearance of real-world entities from text.

The infrastructure for analyzing social big data, built upon our technology, helps us understand situations during disasters such as the 3.11 earthquake and public opinion during the COVID-19 pandemic.

Extremely fast core language technologies

- Self-adaptive classifier that becomes faster when text increases in micrologs



Fastest language analysis beyond machine learning

- Pattern-based morphological analyzer inspired by classification (1M sent./s)

Monitoring the change in real world

- Findings the emergence and disappearance of entities from microblogs

- 【新聞】映画「マトリックスレザレクションズ」12月に公開決定
- Awesome City Club、新曲「Ute still goes on」のリリースを発表
- 【京都水鏡閣】が冬らしい温かな空間に「ふゆあすてくかん」
- Shitabai、VRゲーム「Megamek」などメタバース向け商品3種を発売、パナソニックと共同開発
- 【ゲームパニック秋葉原店(仮称)】アトアーズ秋葉原店後へ2022年冬一斉オープン予定

- 秒速で有名な山崎幸也、今月14日に新店するとの事で食べ納め。
- 宮寺の音楽雑誌「キヤトルム」が廃刊申請へ、関係者より解散を発表したBISHA、デビューから紅白出場までの年を振り返る
- Oculus社より「Oculus Go」のサポート終了の音が公開されました。
- 4月6日に新刊発表、書籍「東京」の巻が公開されました。これに伴い、同日代表取締役を退任いたしました。

Find and structure the change of real world

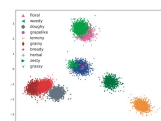
Speak as humans, sensing the unspoken

The meaning of language changes depending on who, when, where, and to whom it is used. Large language models (LLMs), trained on texts from diverse backgrounds, capture the “common denominators” of language and can perform any task. However, adapting LLMs to specific usage contexts remains challenging. We are broadly working on developing **adaptable and portable LLMs** to broaden their applications.

Specifically, we study knowledge-driven analysis, compression, and deconstruction of LLMs; retrieval-augmented generation to supplement their knowledge; multilingual multimodal LLMs with speakers, context (language, space and time), images, and speech; and user simulation using LLMs.

Personalized NLP

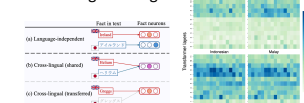
Personalized word vectors in beer reviews



Word meanings vary across speakers

Knowledge Mechanism in LLM

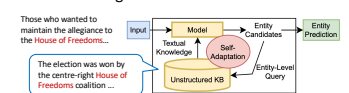
- Inspect neurons for knowledge probing
- knowledge editing



Understand the LLM’s intelligence

Self-adaptive LLM

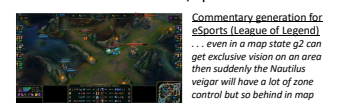
- Complement/adapt LLM’s knowledge
- Retrieval-augmented LLMs



Adapt LLMs to unseen environments

Multilingual Multimodal NLP

- Multilingual corpus w/ vision and audio
- Generation with Vision/Speech LLMs



Understand human perception

Our laboratory works closely with the Toyoda and Goda laboratories, sharing resources such as data, computing, and student space. Members **can pursue interdisciplinary research beyond NLP**, including social informatics using full microblog data, and will receive feedback from diverse perspectives.

Feel free to contact me (ynaga@iis.u-tokyo.ac.jp) for any questions or visit requests.