Linguistic Data for Machine Intelligence

Michael Tseng





Overview Language Technologies

Became One of Surfing's

Most Legendary .

Comedy Central

YouTube - Aug 22, 2019

- "Understanding" and "generation" of textual and spoken language
- Interaction with public or private knowledge bases
- **Dialogue** context
- Multimodality (e.g., language with visual context)

Eddie Aikau						× 🌷 ۹
Q AII	News	▶ Videos	🕼 Images	Shopping	i More	Settings Tools
About 3	850,000 resul	lts (0.52 seco	nds)			
en.wiki	pedia.org > w	iki > Eddie_Aik	au *			
Eddie	Aikau -	Wikipedia				
Edward Hawaii	Ryon Makua an lifeguard a	ahanai Aikau (and surfer. As	Kahului, Hawa the first lifegu	iii, May 4, 1946 - ard at Waimea E	- March 17, 1971 3ay on	8) was a
Favorite waves: Waimea Bay (North Shore, Oa Height: 5 ft 11 in (1.80 m)						
Vears active: 1959-1978 Born: May 4, 1946, Kanului, Territory of Plawali, Life, Memorial surface invitational, Decular sulture, Defenses						erritory of Hawaii,
Life in	errional auto	ing invitationa	r opular cur	ture references		
Peo	ople also	ask				
How did Eddie Aikau die?						~
Wha	at year did B	Eddie Aikau	die?			~
Is E	ddie Aikau	still alive?				~
Wha	t is the sto	ry of Eddie A	Aikau?			~
						Feedback
Video ©	os Fé	DRUNK	2:33	0	3.01	0
Linu	Eddia Aika		Miles Miles	a Eddie Aikers	Linu	iin . The Leaned

And What Happened To

YouTube - May 4, 2019

Him? Google

Dadakoglu

of Eddie Aikau Official

YouTube - Sep 11, 2013

Trailer - 30 for 30

Quiksilver



Edward Ryon Makuahanai Aikau was a Hawaiian lifeguard and surfer. As the first lifeguard at Waimea Bay on the island of Oahu, he saved over 500 people and became famous for surfing the big Hawaiian surf, winning several awards including the 1977 Duke Kahanamoku Invitational Surfing Championship. Wikipedia

Born: May 4, 1946, Kahului, HI Died: March 17, 1978, Hawaii Spouse: Linda Crosswhite (m. 1972-1978)

Lifeguard

Residence: Kahului, Hawaii, United States

Siblings: Clyde Aikau, Myra Aikau

Parents: Sol Aikau, Henrietta Aikau





"Traditional" Paradigm Data Sets for NLP

- Pipeline metaphor (e.g., tokens → morphology → synta x → semantics)
- Formal model of features
- **Strict** labeling schemes with low tolerance for ambiguity



"Traditional" Paradigm Linguistic Annotation

Trained Experts

Usually **linguists** or others well versed in the formal modeling

Technical Discussion

Annotation **guidelines**, **adjudication** meetings, etc.

Complex Tasks & Tools

Interfaces designed for specialists, similar to an **IDE**

Crowdsourcing User-Aware Approaches

- Reframe model-driven tasks and guidelines in terms of **accessible concepts**
- Compose task as sequence of smaller, simpler tasks
- Use sophisticated resolution techniques to synthesize an object not directly annotated

Self-Operating Napkin



Crowdsourcing User-Driven Approaches

- Avoid forcing linguistic abstractions on non-experts
- Treat "the crowd" as experts in everyday language use
- Develop **humanistic criteria**: "Can another person understand this?"
- Embrace **ambiguity and discovery** as part of the process







Example Domain Question Answering

- What **sections** of a document answer a given question? What are the **best-justified answers**?
- How do we collect **interesting** and **realistic** questions?
- How do we model and account for **ambiguity**?

See Kwiatkowski *et al.* (2019) and <u>blog post;</u> Clark *et al.* (2020) and <u>blog post;</u> Min *et al.* (2020)

nperial Palace

e free encyclopedia

'ial Palace (皇居 Kōkyo, literally "Imperial Residence") is the primary residence of e park-like area located in the Chiyoda ward of Tokyo and contains buildings includen), the private residences of the Imperial Family, an archive, museums and adm

te of the old Edo Castle. The total area including the gardens is 1.15 square kilom uring the height of the 1980s Japanese property bubble, the palace grounds were value of all of the real estate in the state of California.^{[2][3]}



Question:

who lives in the imperial palace in tokyo?

Short Answer:

the Imperial Family

Long Answer:

The Tokyo Imperial Palace (皇居, Kōkyo, literally "Imperial Residence ") is the primary residence of the Emperor of Japan. It is a large park - like area located in the Chiyoda ward of Tokyo and contains buildings including the main palace (宫殿, Kyūden), the private residences of the Imperial Family , an archive , museums and administrative offices.

Selected References

- Chang et al. 2016. <u>Linguistic Wisdom from the Crowd</u>. In Crowdsourcing Breakthroughs for Language Technology Applications (HCOMP 2015).
- Clark *et al.* 2020. <u>TyDi QA: A Benchmark for Information-Seeking Question Answering in Typologically Diverse Languages</u>. In *TACL*.
- Kwiatkowski et al. 2019. Natural Questions: A Benchmark for Question Answering Research. In TACL.
- Min et al. 2020. <u>AmbigQA: Answering Ambiguous Open-domain Questions</u>. In Proc. EMNLP.