

Tēnā koutou, tēnā koutou, tēnā koutou katoa

Natural Language Processing for Q&A in Indigenous Languages

Massey University: Huia Jahnke, Hona Black, Ruili Wang, and Feng Hou

The University of Auckland: Michael Witbrock, Ziqi Wang, Joshua Bensemann

Cultureflow: Sarah Reo

National University of Singapore: See-Kiong Ng and Stephane Bressan

Project background



Catalyst: Strategic – New Zealand-Singapore *Data Science Research Programme*

New Zealand team:

- Māori and Indigenous Education expert
- Māori language expert
- Māori language education expert, and
- Scientists in AI and computational linguistics

Singapore:

- Scientists in AI and computational linguistics

Prof Huia Jahnke, Sarah Reo and Dr Hona Black Prof Ruili Wang, Prof Michael Witbrock and Dr Feng Hou



Conversational Q&A System in the Māori language

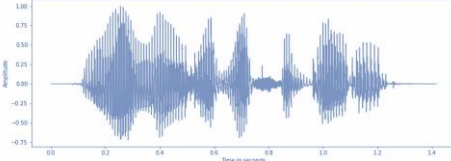
Asking a question in te reo Māori and having the system answer the question in spoken te reo.



Conversational Q&A System for the Māori language



User A



Input

ASR Module

Output

He aha te tāone
matua o
Aotearoa?

Co-led by Prof. Ruili
Wang and Prof.
Michael Witbrock
(NZ)

MT Module

Output

What is the
capital of NZ?

Co-led by Dr Feng
Hou and Prof. Ruili
Wang (NZ)

Q&A Module

Output

Wellington.

Co-led by Prof. See-
Kiong Ng and Prof.
Michael Witbrock
(Singapore & NZ)

MT Module

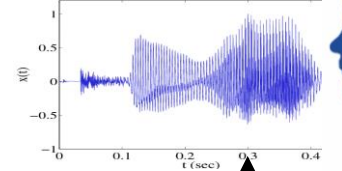
Output

Ko Te Whanganui-
a-Tara.

Co-led by Dr Feng
Hou and Prof. Ruili
Wang (NZ)

TTS Synthesis
Module

Output



User A

Co-led by Prof. Ruili
Wang and Prof.
Michael Witbrock
(NZ)

Four Modules of the Q&A system in Te Reo Māori

- Speech recognition (Professor Ruili Wang, Massey University)
- Speech synthesis/Text-To-Speech (Professor Ruili Wang)
- Machine translation (Dr Feng Hou, Massey University)
- Q&A (Professor Michael Witbrock, University of Auckland)





Singapore team





Singaporean team

- Singapore team focuses on
 - Malay
 - Singlish



Singaporean team

- Singapore team focuses on
 - Malay
 - Singlish





Vision Mātauranga

-- Our recent progress

- Lead by Prof Huia Jahnke (Massey) and Sarah Reo (Cultureflow)
- Partnership
 - We have 8 online meetings with over 40 Iwi leaders
 - Introduce our project
 - Discuss the future of applying AI to te reo Māori
 - Obtain their support
- Participating
 - We have added two new Māori language/culture experts to our team:
 - Dr Jeremy Tātere MacLeod
 - Dr Annemarie Gillies
 - We have 5 Māori Māori language/culture experts in the team and 3 non-Māori team members



Systems under construction

- Speech recognition system
- Speech synthesis system
- Machine translation system
- Q&A system

Position paper

See-Kiong Ng, Ruili Wang, Fangyi Zhu, Feng Hou, Anh Tuan Luu, Huia Jahnke, Michael J. Witbrock, Stéphane Bressan. (2022) **Cross-lingual Question Answering in the Languages of New Zealand and Singapore**, submitted to ACL 2022 theme track (submitted on 16/11/2021) (A* conference)

- Blue indicates the Singaporean team members

Publications

We have submitted 3 journal papers with [Singapore team](#):

- Feng Hou, Ruili Wang, [See-Kiong Ng](#), Michael Witbrock, [Fangyi Zhu](#), Xiaoyun Jia. (2021) “**Exploiting Anonymous Entity Mentions for Named Entity Linking**”. Submitted to *Knowledge-based System* (ERA ranking A).
- Feng Hou, Ruili Wang, [See-Kiong Ng](#), [Fangyi Zhu](#), Michael Witbrock, Steven F. Cahan, Lily Chen, and Xiaoyun Jia. (2021) “**Anisotropic Span Embeddings and the Negative Impact of Higher-Order Inference for Coreference Resolution**”, submitted to *Natural Language Engineering*
- Ruili Wang, Feng Hou, Steven Cahan, Li Chen, Xiaoyun Jia, and Wanting Ji. (2021) **Fine-Grained Entity Typing with a Type Taxonomy: a Systematic Review**. Submitted to *IEEE Transactions on Knowledge and Data Engineering*, second review (ERA ranking A*)

Challenges

- We have difficulty to recruit suitable candidates, especially, Māori PhD students and Post-doc
 - 2 Post-doc positions (4 have been recruited)
 - 4 PhD students (2 have been recruited)
- We are still working on collecting more high-quality speech and language data
 - Finding a common goal and understanding between different teams
 - Sharing data is possible



**CATALYST
FUND**

Advancing global science partnerships
for New Zealand

Catalyst: Strategic – New Zealand-Singapore Data Science Research Programme

KIA ORA TATOU KATOA (thank you)

