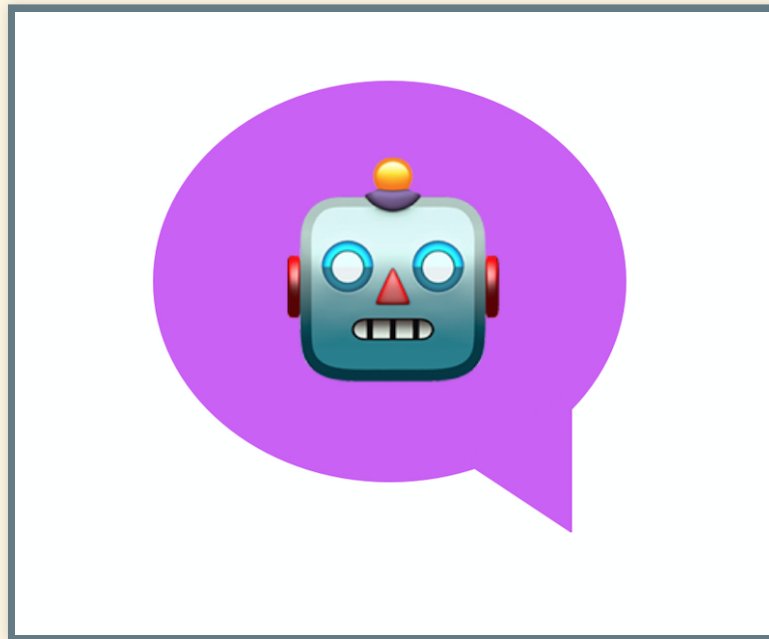


WIKIPEDIA FACT CHATBOT

David Fuhry, Lukas Gehrke, Leonard Haas,
Lucas Schons, Jonas Wolff



INHALT

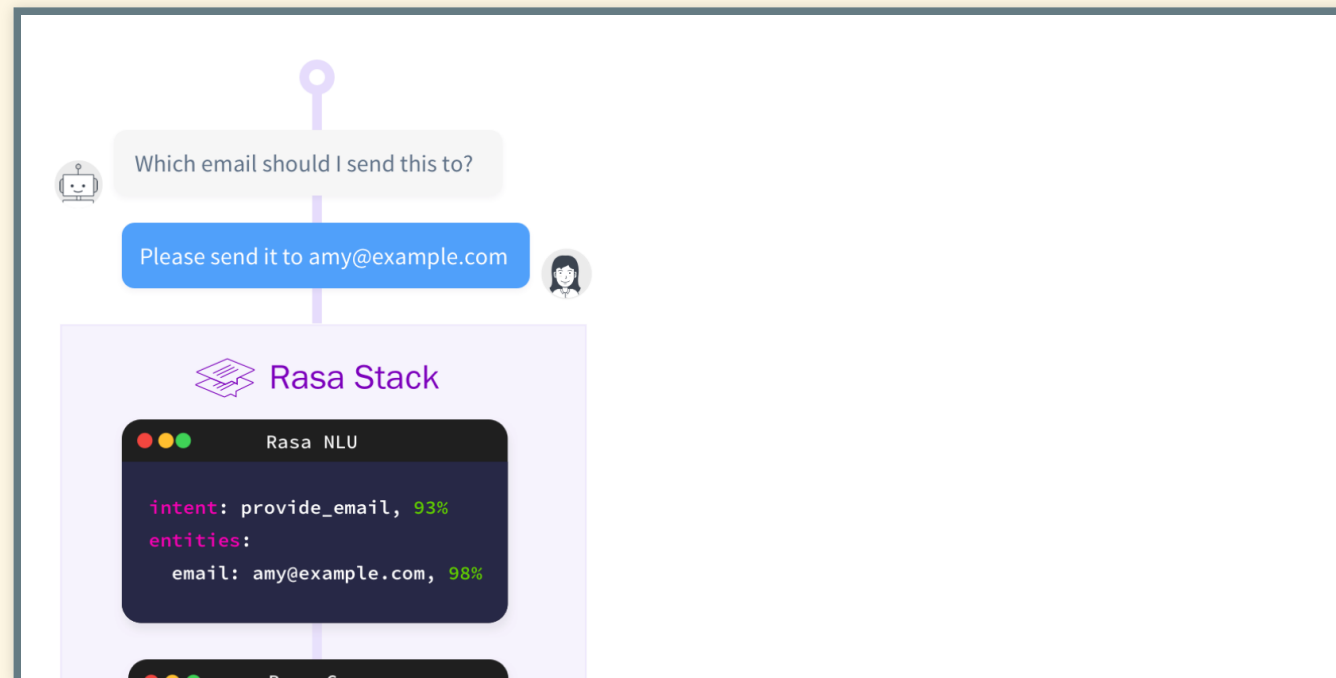
- Thema & Forschungsfrage
- Lösungsansatz
- Ergebnisse

THEMA - RASA CHATBOT

RASA ist ein Framework für die Nutzung von **Chatbots** im Bereich *Conversational AI*. Das System muss mit Facts befüllt werden, um Antworten generieren zu können.

"The Rasa Stack is a set of open source machine learning tools for developers to create contextual chatbots and assistants."

THEMA - ARCHITEKTUR VON RASA



FORSCHUNGSFRAGE

- *Kann man dieses Wissen aus Texten akquirieren?*
- *Kann dies automatisch mit Textquellen durchgeführt werden?*

Dies soll anhand der Wikipedia-Einträge von Physikern erprobt werden

INHALT

- ~~Thema & Forschungsfrage~~
- Lösungsansatz
- Ergebnisse

LÖSUNGSANSATZ - RASA BOT KONFIGURIERERN

- Für Training eines Bots werden *entities*, *text* und *intents* benötigt
- Definition *intent* aus [RASA Docs](#):
"The intent is the intent that should be associated with the text."

LÖSUNGSANSATZ - RASA BOT KONFIGURIERERN

AUFSTELLEN VON INTENTS ZU PHYSIKERN:

- awards
- birthdate
- birthplace
- spouse
- university
- area of research
- ...

LÖSUNGSANSATZ - DATENQUELLE

- 982 englischsprachige Wikipedia-Artikel
- Scraping über jeweiligen Link mit `wikipediR`
- HTML-Formatierung
- Ca. 550 Wörter/Artikel

Igor Kurchatov

From Wikipedia, the free encyclopedia

Igor Vasilyevich Kurchatov (**Russian:** И́горь Васи́льевич Курча́тов; 8(21) January 1903 – 7 February 1960), was a [Soviet nuclear physicist](#) who is widely known as the director of the [Soviet atomic bomb project](#). Along with [Georgiy Flyorov](#) and [Andrei Sakharov](#), Kurchatov is known as the "father of the Soviet atomic bomb" and, later, "the father of the Soviet Nuclear Missile" for his directorial role in a clandestine Soviet nuclear program formed during [World War II](#) in the wake of the USSR's discovery of the [Western Allied efforts to develop nuclear weapons](#). After nine years of covert development, as well as Soviet spies successfully infiltrating the [Manhattan Project](#), the Soviet Union successfully tested its first nuclear weapon, codenamed *First Lightning* at the [Semipalatinsk Test Range](#) in 1949. In 1954 he was awarded the [USSR State Prize](#) in physics.

From 1940 onward, Kurchatov worked on and contributed to the advancement of the nuclear weapons program, and later advocated for the peaceful development of nuclear technology. In 1950, Kurchatov contributed in the development of the [Hydrogen bomb](#) with Andrei Sakharov who originated this development as *Sakharov's Third Idea*. Other projects completed under Kurchatov included the installation and development of Soviet Union's first particle accelerator, the [Cyclotron](#), inauguration and establishment of the



Soviet scientist and physicist Igor Kurchatov

LÖSUNGSANSATZ - PROCESSING

```
|— packages.list
|— script
|   |— master.R
|— wikiproc
|   |— DESCRIPTION
|   |— NAMESPACE
|   |
|   |— R
|       |— get_awards.R
|       |— get_birthdate.R
|       |— get_birthplace.R
|       |— get_data.R
|       |— get_spouse.R
|       |— get_university.R
|   |— man
|   |— tests
```

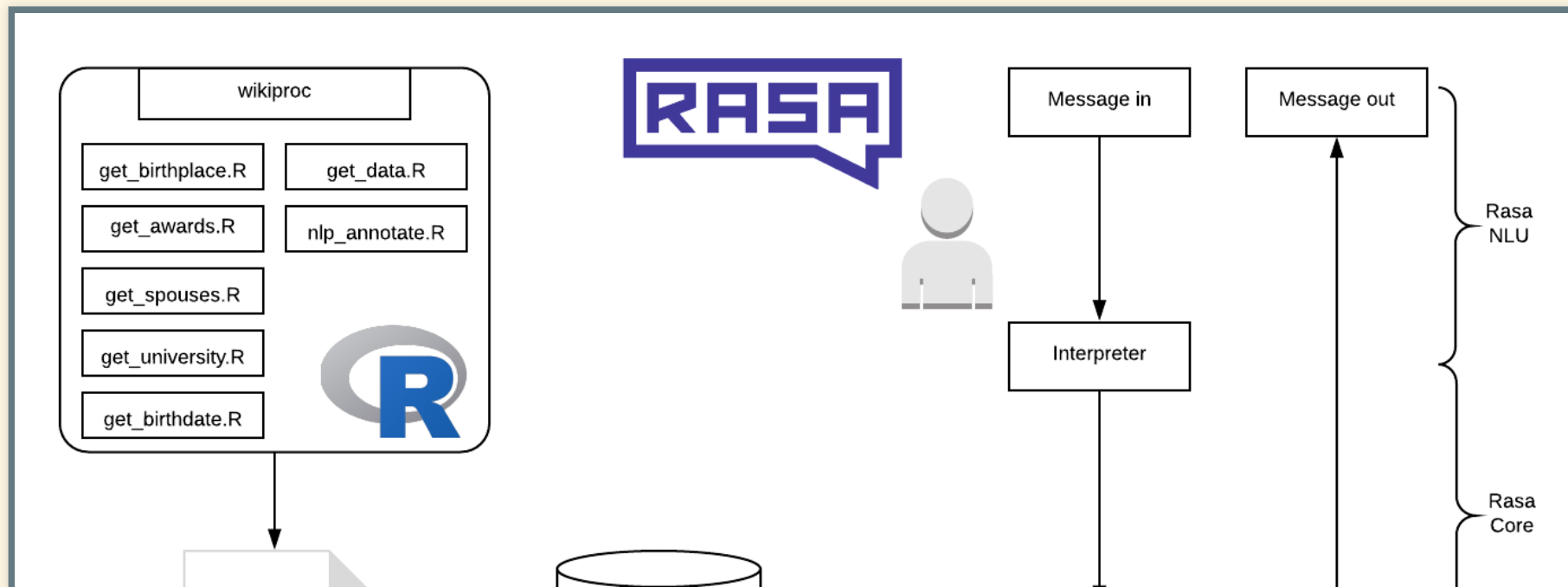
LÖSUNGSANSATZ - PROCESSING, EXTRAKTION INTENTS

- Beispiel `R/get_awards.R`
- Nutze *NER-Entities* und *Regex*

LÖSUNGSANSATZ - PROCESSING, EXTRAKTION INTENTS

- Beispiel `R/get_spouse.R`
- Nutze *Pattern* sowie *NER-Entities* und *POS-Tags*

LÖSUNGSANSATZ - GESAMTARCHITEKTUR



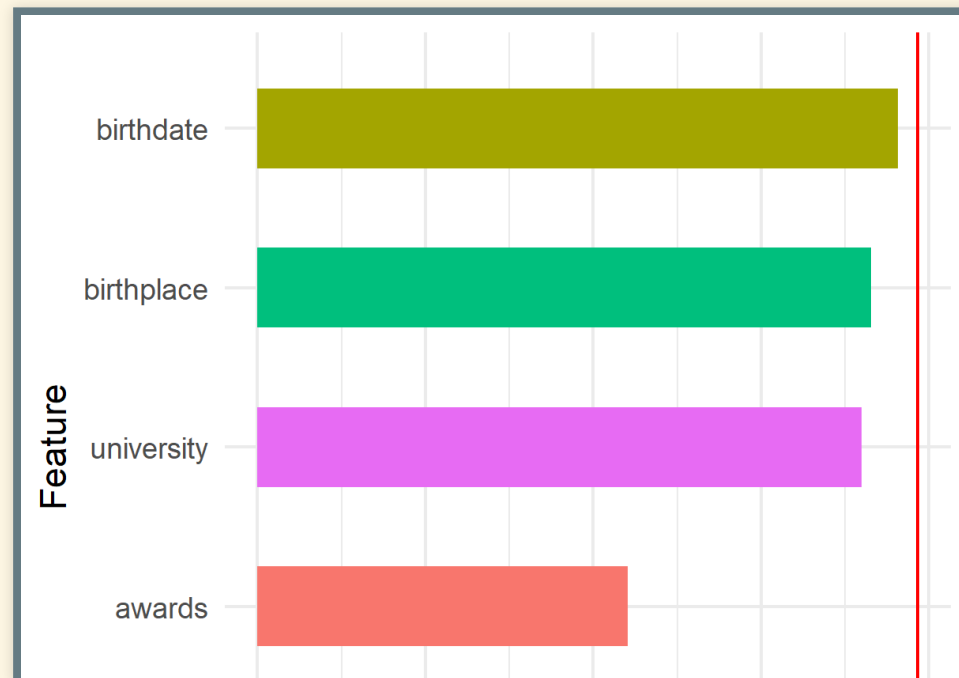
VORFÜHRUNG RASA - BOT

INHALT

- ~~Thema & Forschungsfrage~~
- ~~Lösungsansatz~~
- Ergebnisse

ERGEBNISSE

ANZAHL GEWONNENER ERGEBNISSE ZU INTENTS; REST IST NA



BEWERTUNG SOFTWARE/DATENGRUNDLAGE I

1. RASA-BOT

- (+) NLP funktioniert in Rasa sehr gut
- (+) Frei konfigurierbares Skript (actions.py)
- (-) RASA-Software schwierig aufzusetzen
- (-) RASA wurde kurz vor Projektbeginn stark umgestellt

BEWERTUNG SOFTWARE/DATENGRUNDLAGE II

2. WIKIPEDIA ARTIKEL

- (+) Relativ einheitlicher Aufbau (Einleitung, Werdegang etc.)
- (-) Unterschiedlich ausführlich, tlw. Fehlen von Daten

BEANTWORTUNG DER FORSCHUNGSFRAGE

- Definieren der Intents für den Bot sollte vorab geschehen
- Informationen zu Intents unterschiedlich schwierig aus Text zu extrahieren

ERGEBNISSE

Kann man dieses Wissen aus Texten akquirieren?

Kann man dies automatisch mit Textquellen durchführen?

Es ist möglich, Fakten zu vordefinierten Intents aus Texten zu extrahieren und diese dem Bot zur Verfügung zu stellen.