Software Requirements Specification

Evil Scientists: Caterina Ceccato - Chia Ying Lu - Jos Prinsen - Robyn de Roode - Esther Senneker - Anita Vrins - Emil Anker Wiik

1. Introduction

This document contains all the requirements for the agreed upon software. The following pages outline the user-, system-, and AI-specific requirements.

1.1. Purpose

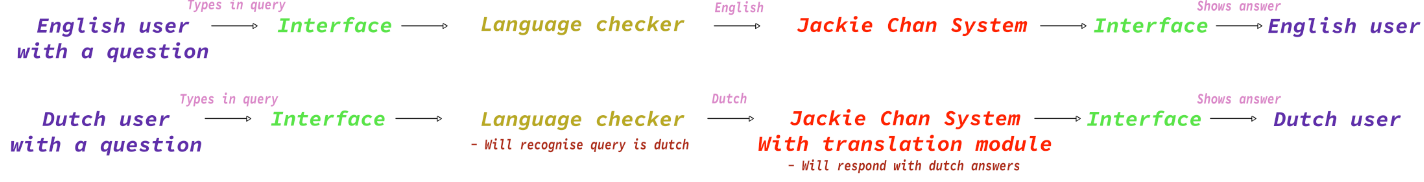
Jackie Chan movies are similar to each other, which makes it difficult to identify the movie you are looking for if one does not remember the title. Our friendly chatbot aims to solve this problem by interacting with the user via questions, and matching the answers with movies in its database.

1.2. Intended audience

This document contains information relevant for the developers, the client Anazito LTD, users, and testers. In this document, we outline the requirements for the software including user-, system-, and AI specific requirements. This document encompasses all relevant information regarding the system and its requirements.

2. User requirements

1. The system should give an error message if it cannot return an output with an expected accuracy above 75%.
2. **Interface:** 
   1. The system shall have a graphical user interface which can take input queries related to Jackie Chan movies.
   2. The user interface should help prevent errors by giving 3 example input queries to the user.
   3. The interface should be structured with an input and output section, showing what is given as input from the user, and what is given as response from our system. It will also show 1 question previous to the current question, both input and given output.
   4. The system will show the output in the GUI in a text box labeled output.
3. **Translation:** 
   1. The system will translate all Dutch inputs to English inputs in order to be properly processed by the Jurassic model.
   2. When a Dutch input is detected, the system will translate the English output from the Jurassic model to Dutch.
   3. The system will respond to the (translated) query by text, using the Jurassic API.

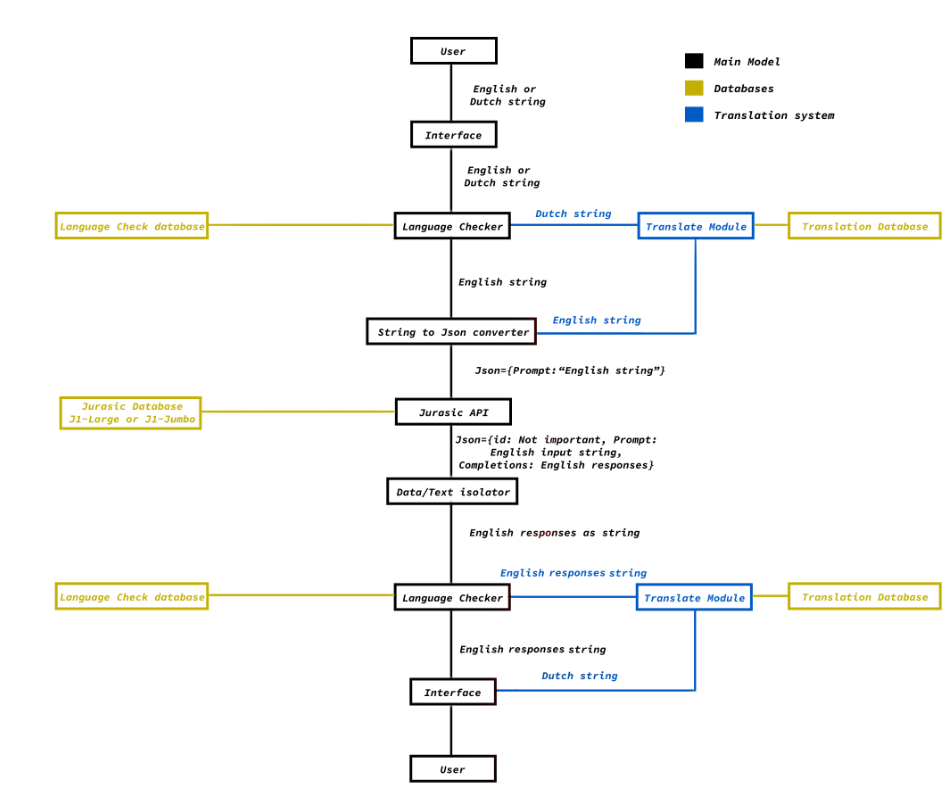


*A diagram showing a high level view of what the Jackie Chan system will do.*

3. System requirements

3.1. Functional requirements

1. **Usability requirements**
2. All users shall have access to the system at an aggregate level of 156 questions per day. The free tier for the Jurassic-1 (178B parameters) model limits daily access to 10K tokens, meaning about 2500 words, or 156 questions of 16 tokens each.
3. The starting input is a question posed by a user.
4. The system’s features and functionality should be described to the users.
5. The English string shall be converted to a JSON object.
6. The system shall use the Jurassic API (custom model: tilburg-jackie-chan-v0.1) to answer the user queries.
7. The system shall recognize Dutch as a language, when given as input by the user and translate the output back in Dutch if the input is in Dutch.
8. **Memory requirement**
   1. Every previous input shall be temporarily remembered by the system.
   2. A user should be able to look at previous input and output.
9. **Speed performance requirements**
10. The system shall give an output in no more than 200 seconds.
11. The translation of inputs and outputs shall take no more than 120 seconds.
12. **Security requirements**
13. A number should be assigned to each user to be able to identify misuse of the system.
14. The system shall not save the data inputted by the user to an external drive.
15. The system shall have a Captcha to prevent bots from overloading the system.



3.2. Non-functional requirements

1. **Product requirement**
   1. The translation model should load into the GPU memory in no more than 120 seconds.
   2. The interface should be a GUI which the user can give inputs to and can read of the outputs given by our Jurassic model.
   3. The system should not experience critical failures.
   4. The product should be available to the user for 1 year.
   5. The product should be easy to use by the user.
   6. The system should be stable, complete and robust.
   7. The system should provide both Dutch and English.
2. **Organizational requirements**
3. **External requirement**
   1. The system shall not disclose any personal information about users
   2. The GUI shall allow users to enter their query in one click.
   3. The system shall respond to the input in less than 120 seconds.
   4. The performance of the system should not be affected by higher workloads.
   5. The system should not include any content that is offensive, insensitive, upsetting, intended to disgust or just plain creepy.
   6. The system should be protected against attacks.
   7. The background colour for all screens should be white.

4. AI specific requirements

1. **The pre-trained Dutch-English translation model**
   1. Should translate Dutch and English queries with above 75% accuracy
   2. Should produce a response within 120 seconds
2. **The pre-trained conversation model**
   1. Should respond accurately to prompts with syntax structure. A custom model should generate accuracy above the 86% baseline reached for a 16-example prompt.
   2. Should respond timely. Custom models with task-specific behaviour can process a typical request in less than 50 milliseconds.
   3. Should respond to select topical queries on production year, release date, IMDB rating, original title, character names, plot/overview, director and slogan/tagline.
   4. Should accept input size up to 64 tokens. At an average 4 tokens per word this limits the input prompt at 16 words. Jurassic-1 models can handle up to 2048 tokens or 512 words.
   5. Should accept the names of lead actors, characters and directors as input to fulfill named queries
3. **Both models**
   1. Will not aggregate or store training data by selection, collection, preprocessing, filtering, categorization, or feature extraction
   2. Will not dynamically improve performance by retraining on new data