Document Management with Formal and Independent Knowledge Representation

NASA Space Apps Challenge 2023

October 8, 2023

Motoshi Horii (Team: SpaceCavies)

Background

- A large number of recommended practices (knowledge) are proposed in space development.

- This knowledge is scattered across various documents.

- Various documents are each added, modified, and deleted.

- In space development, it is important to follow these knowledge.

NASA-HDBK-2203	B	NASA Software Engineering Handbook	2020-04-20	0		
NASA-HDBK-4001	Baseline	Electrical Grounding Architecture for Unmanned Spacecraft	1998-02-17	0		
NASA-HDBK-4002	B	Mitigating In-Space Charging Effects-A Guideline	2022-06-07	0		
NASA-HDBK-4006	A	Low Earth Orbit Spacecraft Charging Design Handbook	2018-11-28	0		
NASA-HDBK-4007	Baseline w/CHANGE 3	Spacecraft High-Voltage Paschen and Corona Design Handbook	2016-10-27	3	2020-12-29	2020-12-29
	Baseline					

Over 80 documents are registered in NASA Standards.
(This is just a small part of the whole.)

Challenge

While AI is convenient, managing text with AI leads to the following problems:

1. Comprehensiveness Issue
2. Protection of Confidential Information

Comprehensiveness Issue

- AI operates on computational platforms, its computational capacity has limits.

- AI may ignore complex commands to huge sentences.

- It's challenging to determine if AI's observations are truly exhaustive.

- Comprehensiveness is important in space development because the slightest error can lead to a major accident.

Protection of Confidential Information

- Information required for space development includes confidential data.

- Controlling the output of LLM is challenging.

- Teaching LLM confidential information may result in potential data leaks.

Approach

1. Formalization of Knowledge

2. Independent Management of Knowledge

Formalization of Knowledge

- Manage knowledge with logical formulas (e.g., propositional logic, computational tree logic, linear temporal logic).

- With existing programs, one can manipulate knowledge or conduct comprehensive knowledge checks (e.g., SAT solvers, SPIN, NuSMV).

Independent Management of Knowledge

- We manage knowledge outside of the LLM model.
- We manipulate knowledge outside the LLM model using existing tools.
- Since there's no need to teach the LLM confidential information, it prevents potential leaks of confidential data.

Implementation

1. Knowledge Extraction
2. Knowledge Formalization
3. Document Review

Knowledge Extraction

- Input the document and extract recommendations in natural language.
- Utilize LLM.

Output (in Japanese)

- "Design burst pressure refers to the calculated pressure that a component should withstand during qualification tests without rupturing or bursting. This value is a recognized figure calculated in an industry and/or government accepted method to determine design pressure."
- "Pressure vessel and system testing should be conducted as a hydrostatic test, which involves filling the vessel or system with a liquid (usually water) and applying pressure up to a specified level."
- "To ensure safety and prevent propagating explosive effects to other elements, fuses must have an interrupt mechanism."

Output (in English)

Knowledge Formalization

- Formalizing natural language using LLM
- Expressing propositional logic formulas in S-expressions (formula)
- Maintaining a list of each atomic term and its description (atoms)

```
{
  "prop": "A specific load project should create a mission-specific draft version o
f this document. The purpose is to eliminate all non-applicable requirements so that
only applicable requirements remain after detailed adjustments and waiting for the re
view of the PSWG and Range Safety.",
  "atoms": [
    {
      "id": "MissionSpecificDraft",
      "description": "A mission-specific draft is created"
    },
    {
      "id": "NonApplicableRequirementsEliminated",
      "description": "Non-applicable requirements are eliminated"
    },
    {
      "id": "WaitforPSWGRangeSafetyReview",
      "description": "Wait for PSWG and Range Safety review"
    }
  ],
  "formula": "(imply MissionSpecificDraft (and NonApplicableRequirementsEliminated
WaitforPSWGRangeSafetyReview))"
},
```

Document Review

- Bottom-up approach
  - Identify atomic terms with similar meanings and ask humans.
  - If the meanings are similar, ask humans whether the propositional logic formulas constructed from those atomic terms hold true.
  - If those propositional logic formulas are valid, then they should be added to the document.

```
{
  "id_pair": [
    "RiskMitigation",
    "RiskManagement"
  ],
  "reason": "Both are involved in the process of planning and controlling for potential risks."
},
{
  "id_pair": [
    "RiskAssessment",
    "RiskAnalysis"
  ],
  "reason": "Both activities involve the identification and investigation of potential risks."
},
{
  "id_pair": [
    "SoftwareAssurance",
    "ProductVerification"
  ],
  "reason": "Both refer to processes that ensure the reliability and safety of software."
},
{
```

Document Review

- Topdown approach
    - Ask if a similar statement holds true.
    - If it does, then its content should be added to the document.

```
--->
[Question]
Are these properties same ?

(Property in KDB)
All pressure systems should have a pressure detection device that is marked to indicate the upper an
d lower limits of system pressure.

(Property in TARGET)
Inhibit is an independent and verifiable mechanical or electrical device that prevents the occurrenc
e of dangerous events and should have direct control.

[Answer]
=> Yes
<---
```
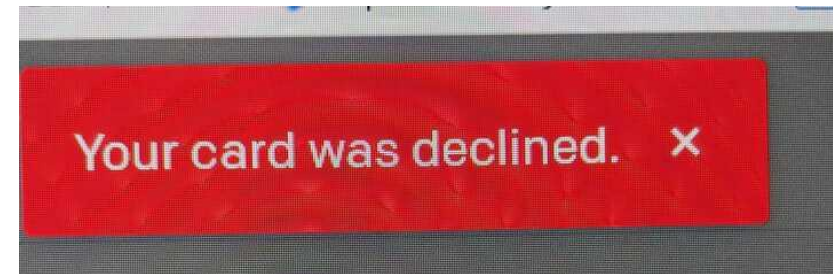
Conclusion

- We have proposed a new document management method that manages and manipulates knowledge outside of the LLM.

- This allows for comprehensive knowledge checks and addresses potential leaks of confidential information.

Thanks!

Your card was declined. ✕

Since my credit card was rejected by OpenAI and not available, much of the processing was implemented using scikit-learn instead of AI.Because of this, the accuracy is a bit poor.