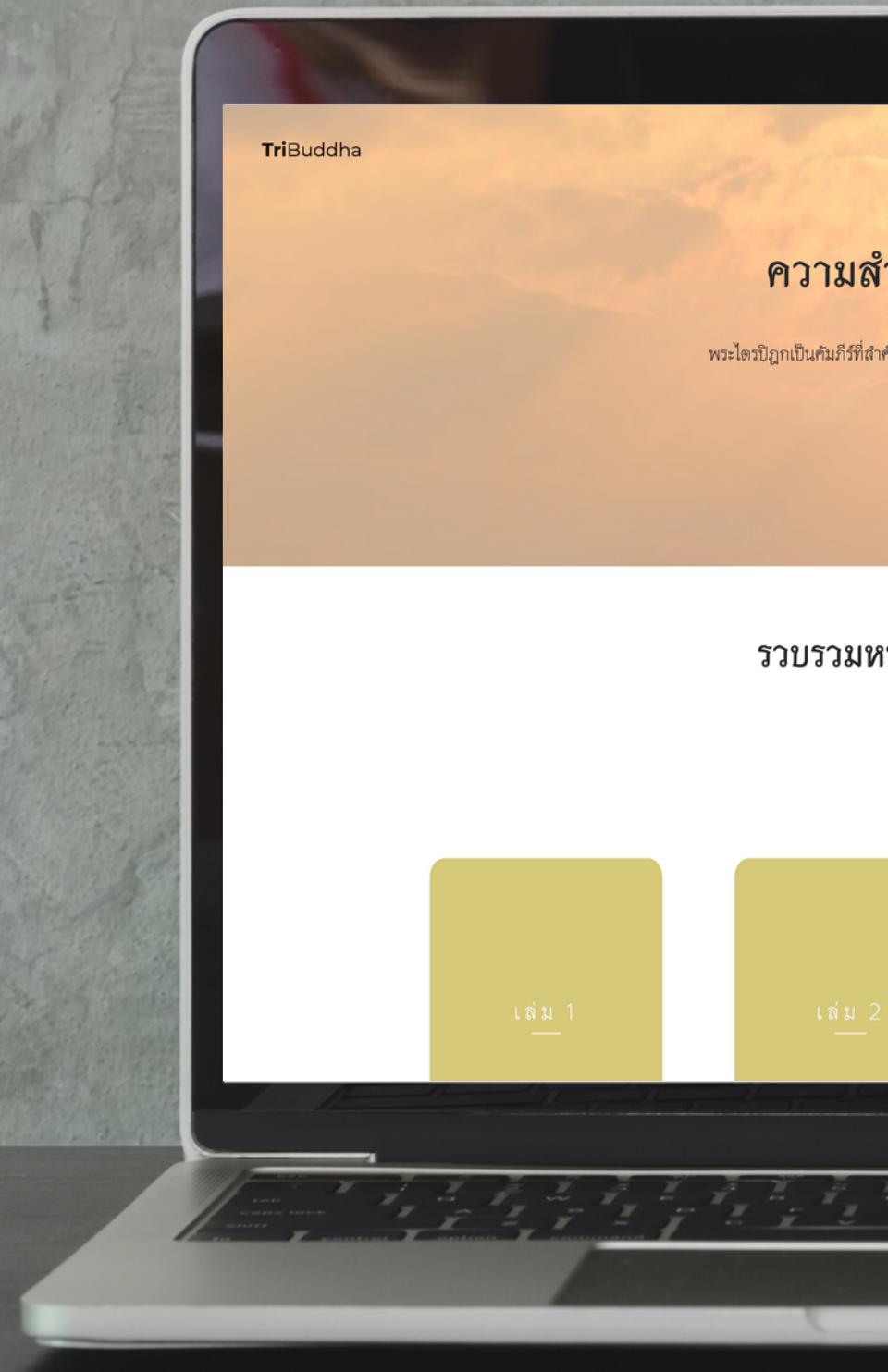


SENIOR PROJECT

TRIBUDDHA

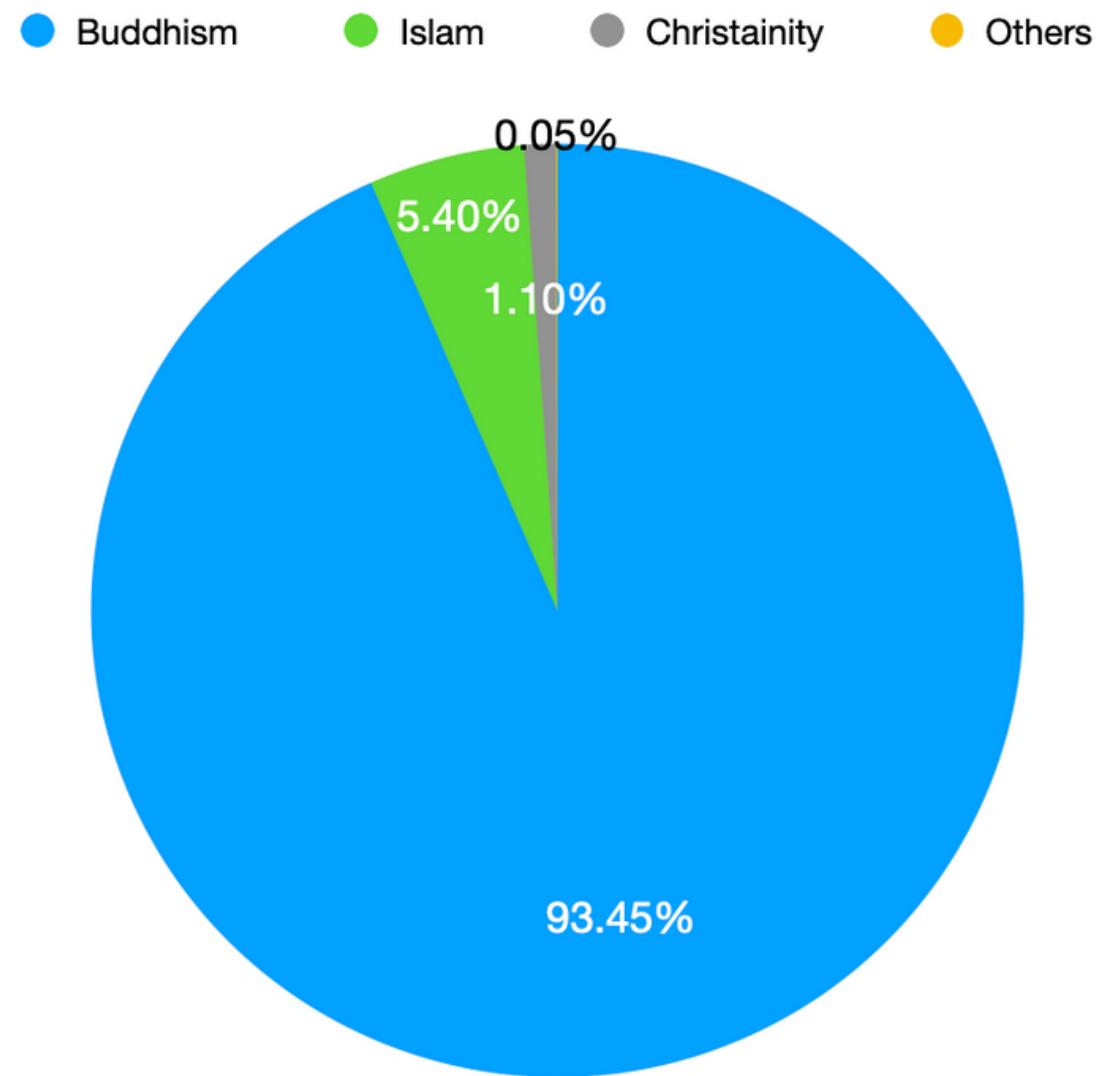
The web application compiles 45 Tripitaka books and along with a fast search system and additional features.

PAGE 1



Motivation

Thailand was 69 million, and 93.45% of Thai population was Theravada Buddhist.



Educational disparities



Problems

01 Tripitaka textbook

- Tripitaka books are expensive.
- Tripitaka books are difficult to access and disseminate

02 Existing Tripitaka websites

- Has no search functionality or find specific content is difficult
- Has no ability to rank the result by relevant content
- Has no ability to compare with Thai - Pali version.

Objective

To develop an online Tripitaka website with functionalities that support users easily accessible Tripitaka contents.

Scope



- Tripitaka web-based application
- TriBuddha collects data of 2 versions including,
ລັບໜາວ & ລັບບາລີ ອັກຊະໄທຍ
- Thai language user interface.

Target users

Thai monks

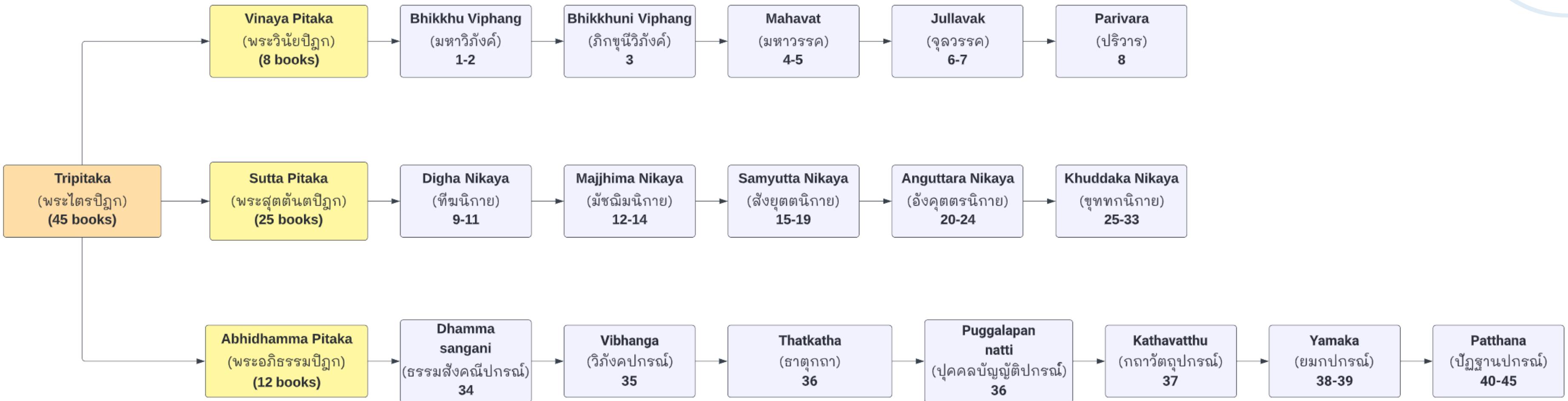
- Have experience in using web browser
- Want to study Tripitaka
 - e.g., prepare for dhamma 6-9 (ເປົ້າຍະສັນ 6-9)
- Need a reliable reference of Tripitaka



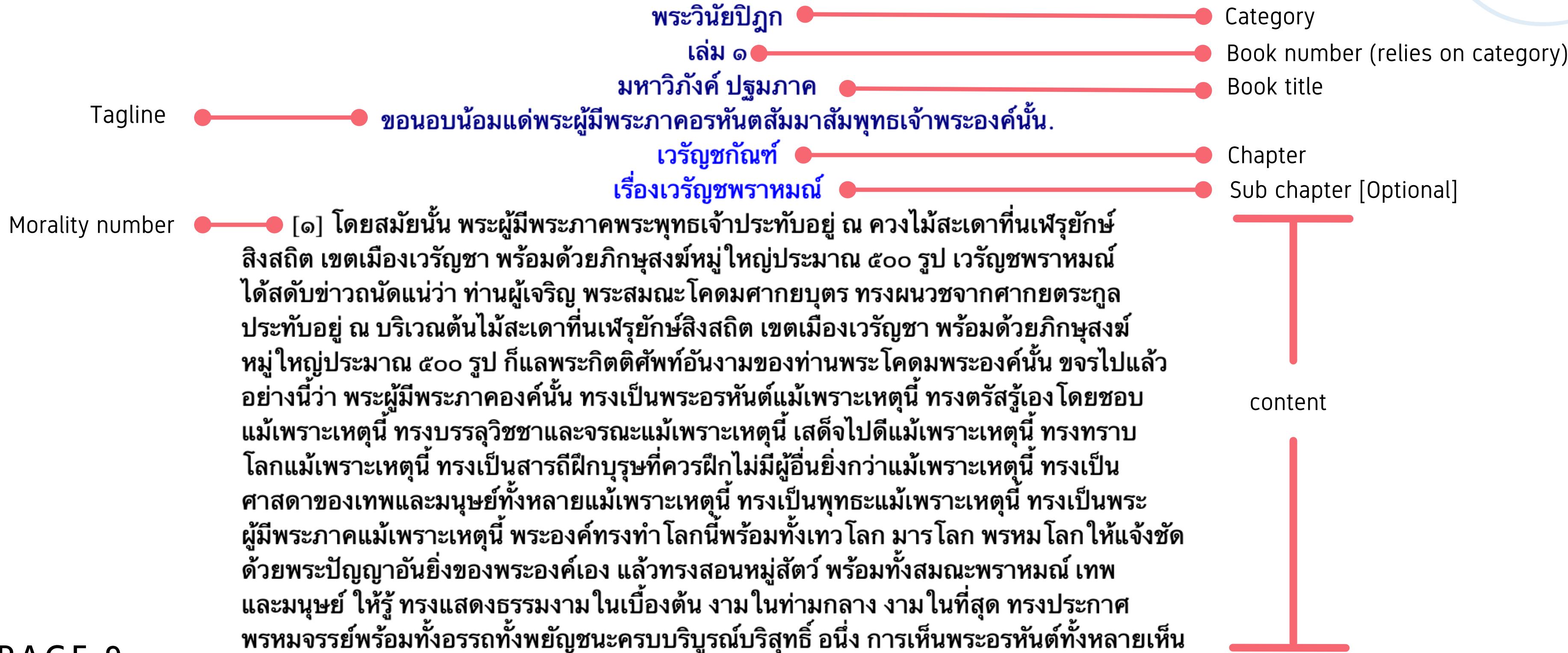
Tripitaka Overview

- A collection of Buddhist teachings of the Theravada (ເກຣວາທ) Buddhist philosophy.
- Structure of Tripitaka is divided into 3 categories such as Vinanya pitaka (ພຣະວັນຍີປຶກ), Sutta pitaka (ພຣະສຸຕຕັບຕຶກ), and Abhidhamma pitaka (ອກົດຮຽມປຶກ)
- There are various versions and languages
 - Thai languages such as ຈິບັບຫລວງ, ຈິບັບມາຈຸພ້າຍ, ຈິບັບມານຄຸນຍ
 - Pali languages such as Pali-Thai, Pali-Roman, ຈິບັບສຍາມຮັບ
- The content of Tripitaka is static, but there might be added new versions in the future.

Tripitaka Structure



Tripitaka Structure



Competitor Analysis

There are 5 web applications that we have surveyed consisting of:

Has **no** search functionality

1. Tipitaka.com
2. ພຣະໄຕປົກກອບໄລນ໌
3. ThaiTripidok

Has a search functionality

1. E-Tipitaka
2. 84000.org

Competitor Analysis

Tipitaka.com Homepage

The screenshot shows the homepage of Tipitaka.com. At the top is a painting of the Buddha sitting under a tree with many弟子 (disciples) around him. Below the painting is a quote in Thai: "ดูกรภิกขุทั้งหลาย บัดนี้ เราขอเตือนพากเบร์ว่า สังฆารหั้งหลายมีความเลื่อมไปเป็นธรรมชาติ พากเบร์จะยังความไม่ประมาทให้ถึงพร้อมเดินฯ" followed by three five-pointed stars.

Below the quote is a section titled "สพุทธานั่น ဓมมathan ชินาติ" which includes a logo for "พระไตรปิฎก ONLINE" and "พระไตรปิฎก ออนไลน์ TIPITAKA ONLINE". It also contains text about the purpose of the site: "เว็บนี้จัดทำขึ้นเพื่อเผยแพร่พระธรรมวินัยของสมเด็จพระสัมมาสัมพุทธเจ้า อันมีความงามในเมืองต้น งามในท่านกลาง และงามในที่สุด (ข้อมูลจาก พระไตรปิฎก ฉบับ CD ROM)" and "ขอขอบคุณท่านที่เข้ามาอ่าน ขอให้ทุกท่านมีความเจริญในธรรมยิ่ง ๆ ขึ้นไปเทอญ".

At the bottom, there is a section titled "รายชื่อพระไตรปิฎก" (List of Tripitaka) with subtext "(กรุณาอ่านค่าอ่อนไหวการใช้งานนี้ก่อน)". It states: "ประกาศ : ขณะนี้ไฟล์พระไตรปิฎกแบบ txt จด ใช้งานชั่วคราว เพราะมีรูปแบบเดียวกับไฟล์ htm". Below this is a table showing four entries:

(๑)	พระไตรปิฎกเล่มที่ ๑ วินัยปิฎกที่ ๑ มหาวิวัฒน์ ปฐมภาค	HTM	TXT	ZIP
(๒)	พระไตรปิฎกเล่มที่ ๒ วินัยปิฎกที่ ๒ มหาวิวัฒน์ ทุติยาภาค	HTM	TXT	ZIP
(๓)	พระไตรปิฎกเล่มที่ ๓ วินัยปิฎกที่ ๓ กิจธนวิวัฒน์	HTM	TXT	ZIP
(๔)	พระไตรปิฎกเล่มที่ ๔ วินัยปิฎกที่ ๔ มหาวรรค ภาค ๑	HTM	TXT	ZIP

Pros

- Allow users to download Tripitaka content.

Cons

- Has no search functionality
- Has only 1 of Tripitaka version

Competitor Analysis

พระไตรปิฎกออนไลน์ Homepage

สารบัญเล่มพระไตรปิฎก

พระวินัยปิฎก

- พระไตรปิฎกเล่มที่ ๐๑ วินัยปิฎกที่ ๐๑ มหาวิถังค์ ภาค ๑ [กลับด้านบน]
- พระไตรปิฎกเล่มที่ ๐๒ วินัยปิฎกที่ ๐๒ มหาวิถังค์ ภาค ๒
- พระไตรปิฎกเล่มที่ ๐๓ วินัยปิฎกที่ ๐๓ ภิกขุนิวิถังค์
- พระไตรปิฎกเล่มที่ ๐๔ วินัยปิฎกที่ ๐๔ มหาวรรค ภาค ๑
- พระไตรปิฎกเล่มที่ ๐๕ วินัยปิฎกที่ ๐๕ มหาวรรค ภาค ๒
- พระไตรปิฎกเล่มที่ ๐๖ วินัยปิฎกที่ ๐๖ จุลวรรค ภาค ๑
- พระไตรปิฎกเล่มที่ ๐๗ วินัยปิฎกที่ ๐๗ จุลวรรค ภาค ๒
- พระไตรปิฎกเล่มที่ ๐๘ วินัยปิฎกที่ ๐๘ ปริวาร

พระสูตรดันตปิฎก

- พระไตรปิฎกเล่มที่ ๐๙ สูตรดันตปิฎกที่ ๐๑ ทีชนิกาย สลับขั้นชวรรค [กลับด้านบน]
- พระไตรปิฎกเล่มที่ ๑๐ สูตรดันตปิฎกที่ ๐๒ ทีชนิกาย มหาวรรค
- พระไตรปิฎกเล่มที่ ๑๑ สูตรดันตปิฎกที่ ๐๓ ทีชนิกาย ปาฏิกวรรณ
- พระไตรปิฎกเล่มที่ ๑๒ สูตรดันตปิฎกที่ ๐๔ มัชถินนิกาย มูลปัณณาสก
- พระไตรปิฎกเล่มที่ ๑๓ สูตรดันตปิฎกที่ ๐๕ มัชถินนิกาย มัชถินปัณณาสก
- พระไตรปิฎกเล่มที่ ๑๔ สูตรดันตปิฎกที่ ๐๖ มัชถินนิกาย อุปริปัณณาสก
- พระไตรปิฎกเล่มที่ ๑๕ สูตรดันตปิฎกที่ ๐๗ สังยุตตน

Like | [สารบัญ](#) | [↑ ก](#) | [ก ↑](#) | [กลับด้านบน](#)

Pros

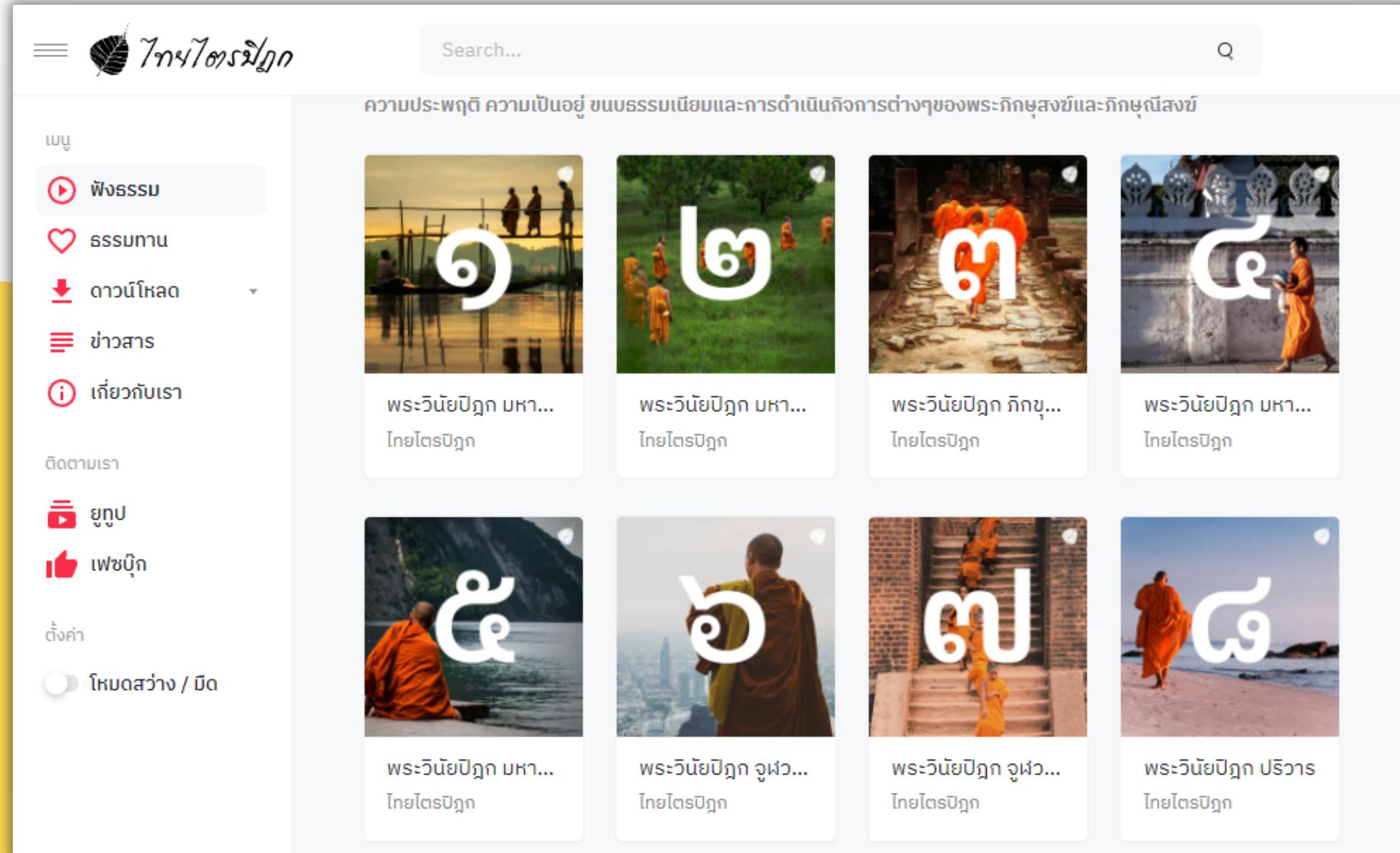
- Allow users to download Tripitaka content

Cons

- Has no search functionality
- Has only 1 of Tripitaka version

Competitor Analysis

Thaitripidok Homepage



Pros

- Allow users to download Tripitaka content
- Has user friendly interface

Cons

- No search functionality
- Has only 1 of Tripitaka version

Competitor Analysis

E-Tipitaka search page

The screenshot shows a search interface with a header containing the E-Tipitaka logo, a search bar, and language selection buttons for 'ไทย (ฉบับหลวง)' and 'อ่านห้องหมอด'. Below the header, there are two search results displayed in columns:

ตรวจหา

"กิกขุทั้งหลาย ! พากกิกขุบริษัทในการ斯นี,
สุดตันตะเหล่าได ที่กีดังขึ้นใหม เป็นค่าร้อย
กรองประเพกษาพย์กลอน มีอักษรرسلลวยมี
พยัญชนะอันวิจิตร เป็นเรื่องนอกแนว เป็นค่า
กล่าวของสาวก เมื่อผู้นำสุดตันตะเหล่านี้มาก
ล่าวอยุ เอาจักไม่ฟังด้วยดี ไม่เชียฟัง ไม่ตั้งจิต
เพื่อจะรู้ทั้วถึง และจักไม่สำคัญว่าเป็นสิ่งที่ตน
ควรศึกษาเล่าเรียน"

"กิกขุทั้งหลาย ! สวนสุดตันตะเหล่าไดที่
เป็นคำของตลาด เป็นข้อความลึก มีความ
หมายซึ้ง เป็นชั้นโลกตระ ว่าเฉพาะด้วย
เรื่องสูญญาตา, เมื่อผู้นำสุดตันตะเหล่านี้
มากล่าวอยุ; เอาจ่ออมฟังด้วยดี ย่ออมเชียฟัง
ย่ออมตั้งจิตเพื่อจะรู้ทั้วถึง และย่ออมสำคัญว่า
เป็นสิ่งที่ตนควรศึกษาเล่าเรียน จึงพากันเล่า
เรียนได้ตาม หวานความแกกันและกันอยุว่า "ขอหนี
เป็นอย่างไร? มีความหมายกี่นัย?" ดังนี. ด้วยการ
ทำดังนี เอาจ่ออมเปิดธรรมที่ถูกปิดไว้ได. ธรรมที่
ยังไม่ปรากฏเอกสาร์ทำให้ปรากฏได, ความสังสัย
ในธรรมหลายประการที่นำสังสัยเอกสาร์บรรเทาลง

เทียบเคียง

"กิกขุทั้งหลาย ! ถ้ากิกขุในธรรมวินัยนี้
กล่าวอย่างนี้ว่า
...(๑) ข้าพเจ้าได้สั่งรับมาเฉพาะพระพักตร์พระ
ผู้มีพระภาค
...(๒) ในอาวสขื่อโน้น มีสังฆอยู่พร้อมด้วยพระ
เกระ พร้อมด้วยปานโภค ข้าพเจ้าได้สั่งมา
เฉพาะหน้าสังฆนั้น
...(๓) ในอาวสขื่อโน้น มีกิกขุผู้เป็นเครืออยู่
จำนวนมากเป็นพหุสูตร เรียนค่ากีร์ ทรงธรรม
ทรงวินัย ทรงมาติกา ข้าพเจ้าได้สั่งมาเฉพาะ
หน้าพระเครือเหล่านั้น
...(๔) ในอาวสขื่อโน้น มีกิกขุผู้เป็นเครืออยู่รูป
หนึ่งเป็นพหุสูตร เรียนค่ากีร์ ทรงธรรม ทรงวินัย
ทรงมาติกา ข้าพเจ้าได้สั่งมาเฉพาะหน้าพระเครือ
รูปนั้น
ว่า "นี้เป็นธรรม นี้เป็นวินัย นี้เป็นค่าสอนของพระ
ศาสนา..."

เรอทั้งหลายยังไม่พึงชื่นชม ยังไม่พึง
ดัดคันคำกล่าวนั้น พึงเรียนบทและพยัญชนะ
เหล่านั้นให้ดี และพึงสอบสวนลงในพระสูตร

Pros

- Has many Tripitaka versions
- Search with various kind of keywords
- Provide filtering content with various criteria
- Has a user-friendly interface
- Compare between 2 Tripitaka versions

Cons

- Has a slow search responses.

Competitor Analysis

84000.org Homepage

The screenshot shows the homepage of 84000.org. On the left, there's a sidebar with links to various Tripitaka versions. The main area has a search form titled "ค้นพระไตรปิฎก ๔๕ เล่ม". It includes fields for selecting a version (พระไตรปิฎก ๔๕ เล่ม ๙ สุต. ๐๑ ที่มนิการ สีลขันธวรค), a date (12/10/19), and a search term (พระไตรปิฎก ๔๕ เล่ม). Below these are sections for filtering by language (ฉบับภาษาจีน อักษรไทย, ฉบับมหาจุฬา), search mode (ค้นทั้งหมด, ค้นเฉพาะพระไตรปิฎกและเชิงอรรถ, ค้นเฉพาะเชิงอรรถ), and a mode selection dropdown (Mode: 1, Start: Book 1 to Book 45). A "RUN" button is at the bottom right.

Pros

- Has many Tripitaka versions
- Search with various kind of keywords
- Provide filtering content with various criteria
- Hit hightlight search keyword(s)

Cons

- Has a user-unfriendly interface
- Has no ablility to compare between 2 versions of Tripitaka.

Competitor Analysis table

	Web Application					
	Tipitaka	พระไตรปิฎก ออนไลน์	Thaitripidok	E-Tipitaka	84000.org	TriBuddha
Search system for Tripitaka	✗	✗	✗	✓	✓	✓
User friendly	✗	✗	✓	✓	✗	✓
Clearly divide 3 categories	✓	✓	✓	✓	✓	✓
Compare multiple versions	✗	✗	✗	✓	✗	✓
Tripitaka E-book	✓	✓	✗	✓	✓	✓

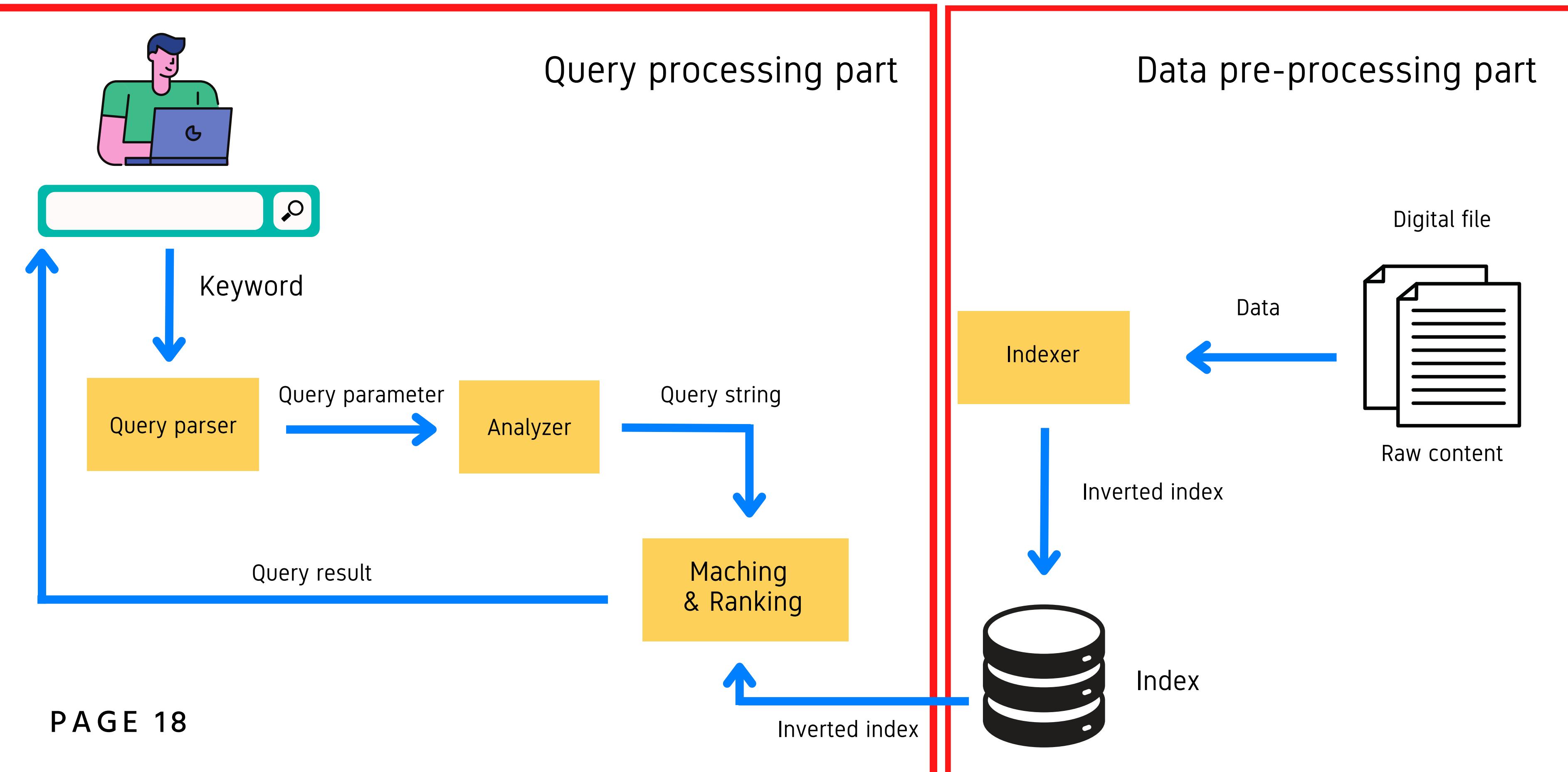
Competitor Analysis table

Search system parts

- ✓ Has
- ✗ None
- Has with unstable results or partially support

	Web Application		
	E-Tipitaka	84000.org	TriBuddha
Simple search (Single words)	○	✓	✓
Advance search (Combination & Concatenate words)	○	○	✓
Filter	✓	✓	✓
Ranking score by relevant content	✗	✗	✓
Highlighting word	○	✓	✓
Query time in 5 seconds	○	✓	✓

Full text search overview



**TECHNICAL
TOOL FOR
SEARCHING**

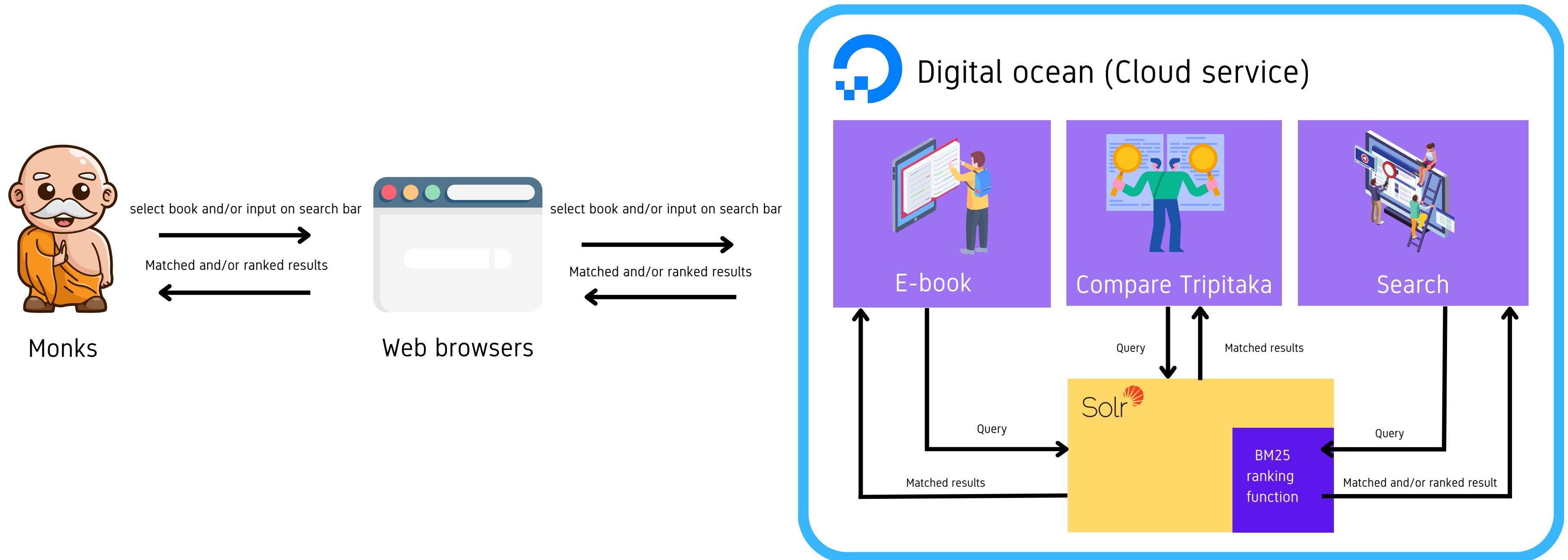
Apache Solr



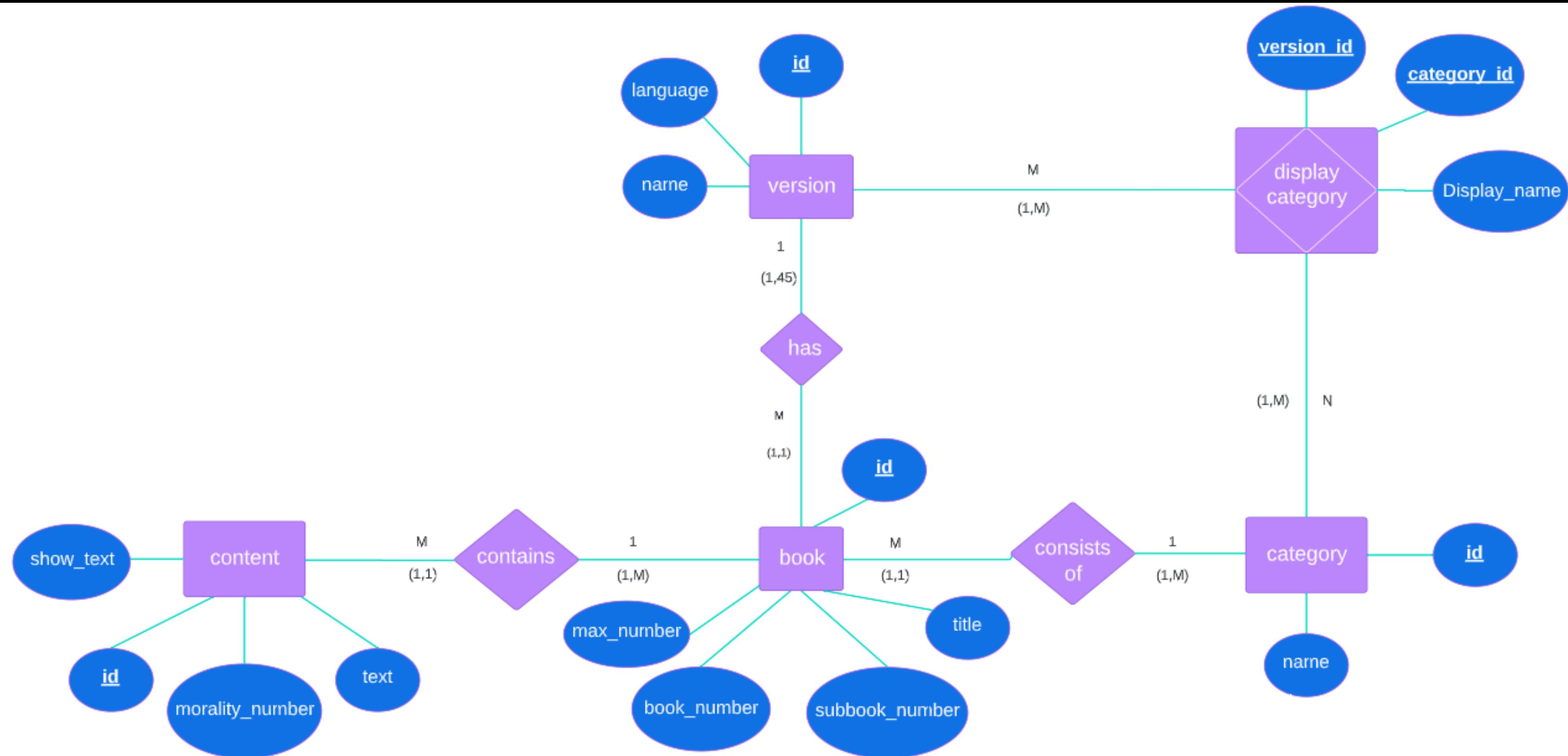
WHY WE USE APACHE SOLR

- Powerful Full-Text Search Capabilities
- Has text analysis to deal with Thai language such as tokenizer, stop words.
- Has hit highlighting functionality that match with query search.
- Has very nice faceting features. For example, in keyword search, you'd break down results by category.
- Comprehensive Administration Interfaces.

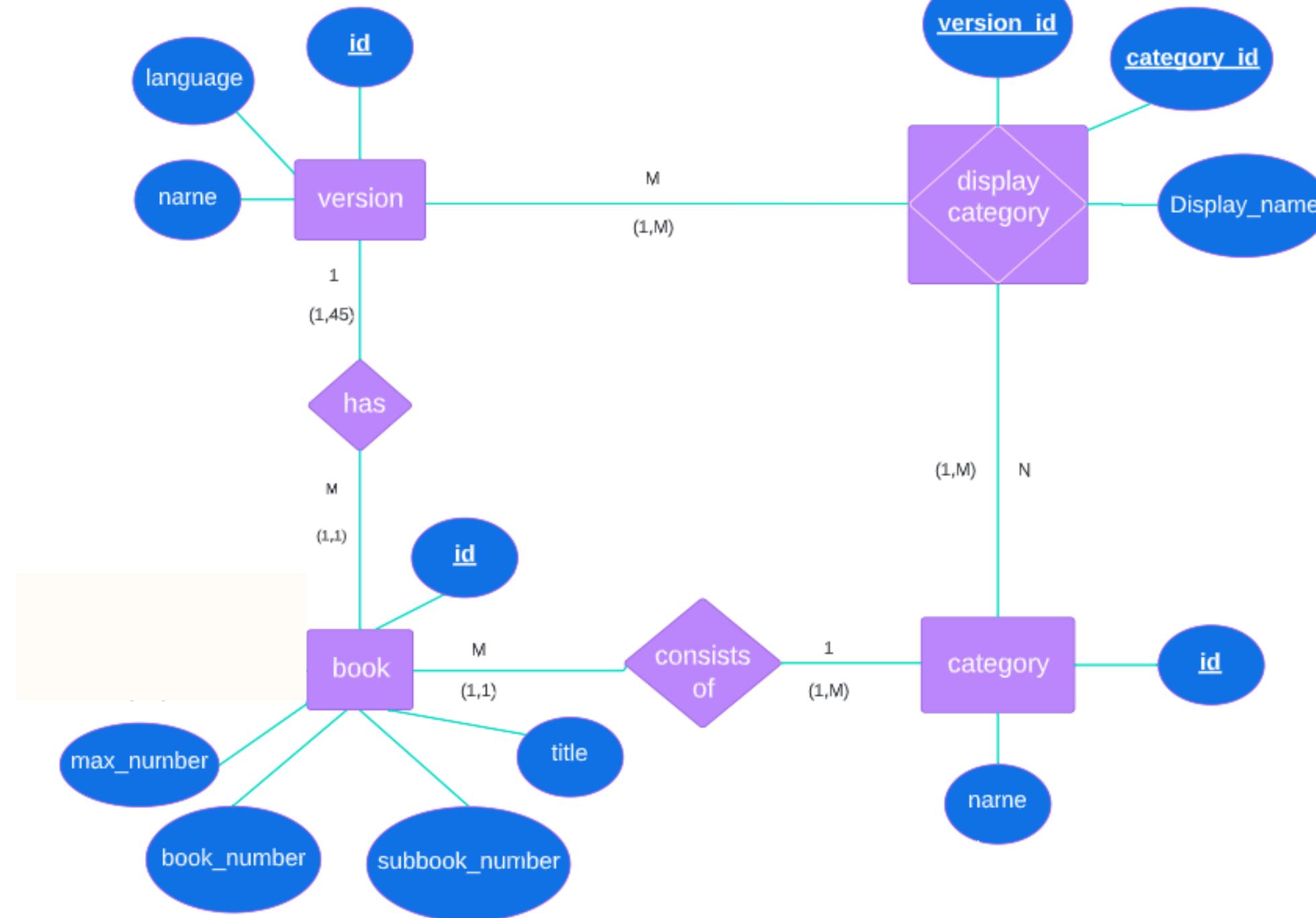
System architecture



ERD



NoSQL schema



```

book_information
{
    "version_id": string,
    "version": string,
    "category": [
        {
            "id": string,
            "display_name": string,
            "book": [
                {
                    "id": string,
                    "title": string,
                    "sub_book": [
                        {
                            "id": string,
                            "max_number": pints
                        }
                    ]
                }
            ]
        }
    ]
}

```

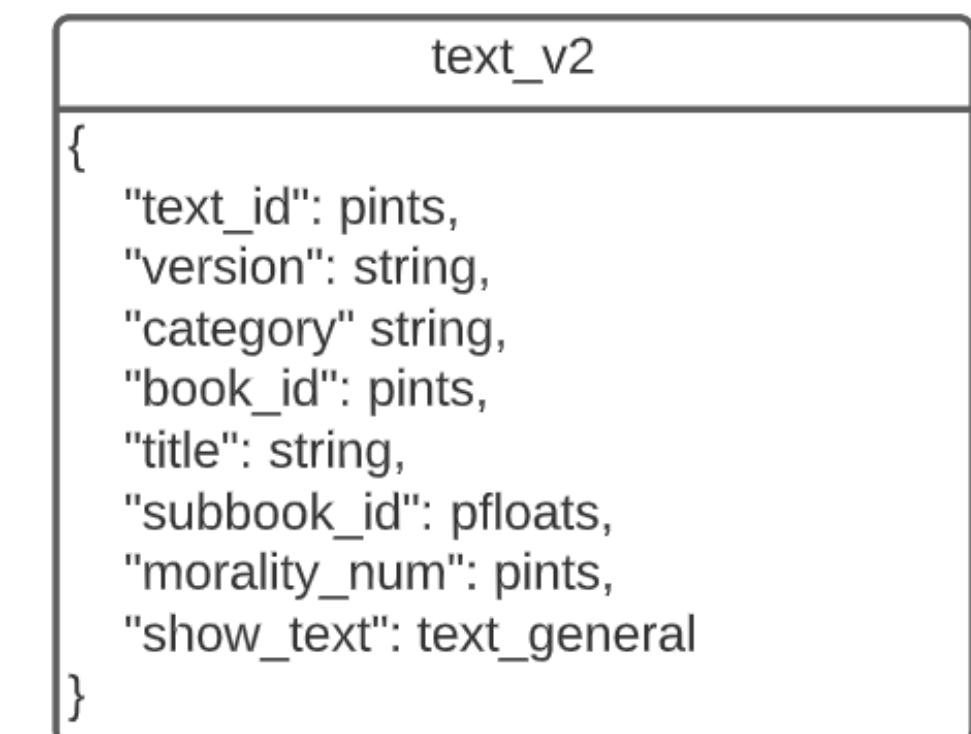
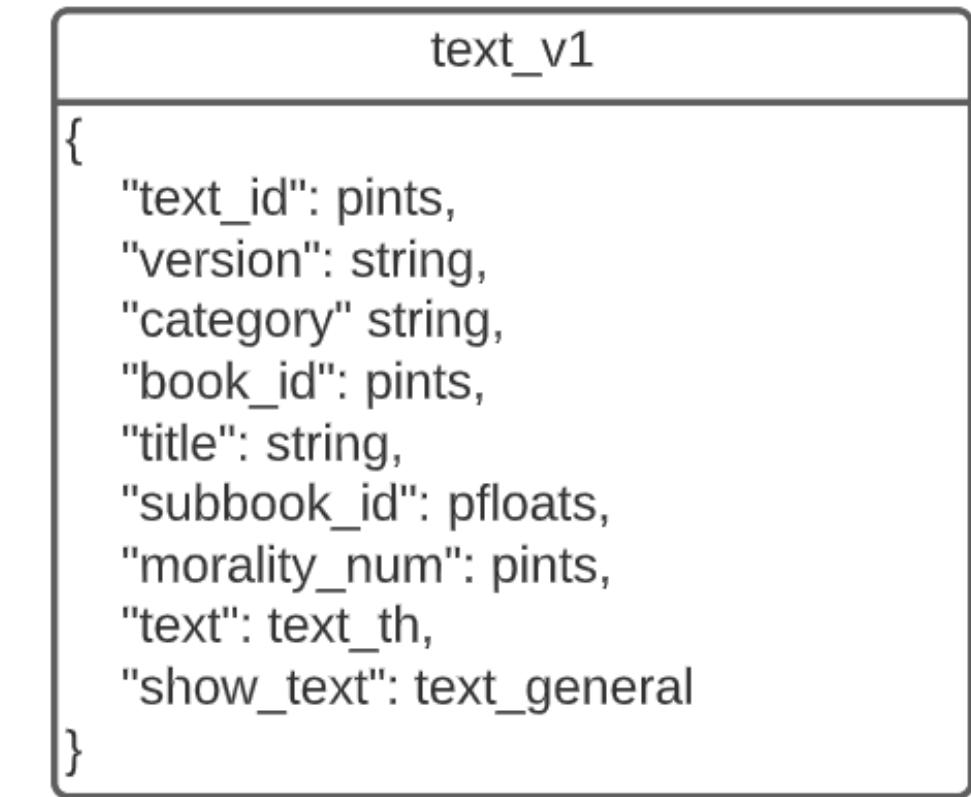
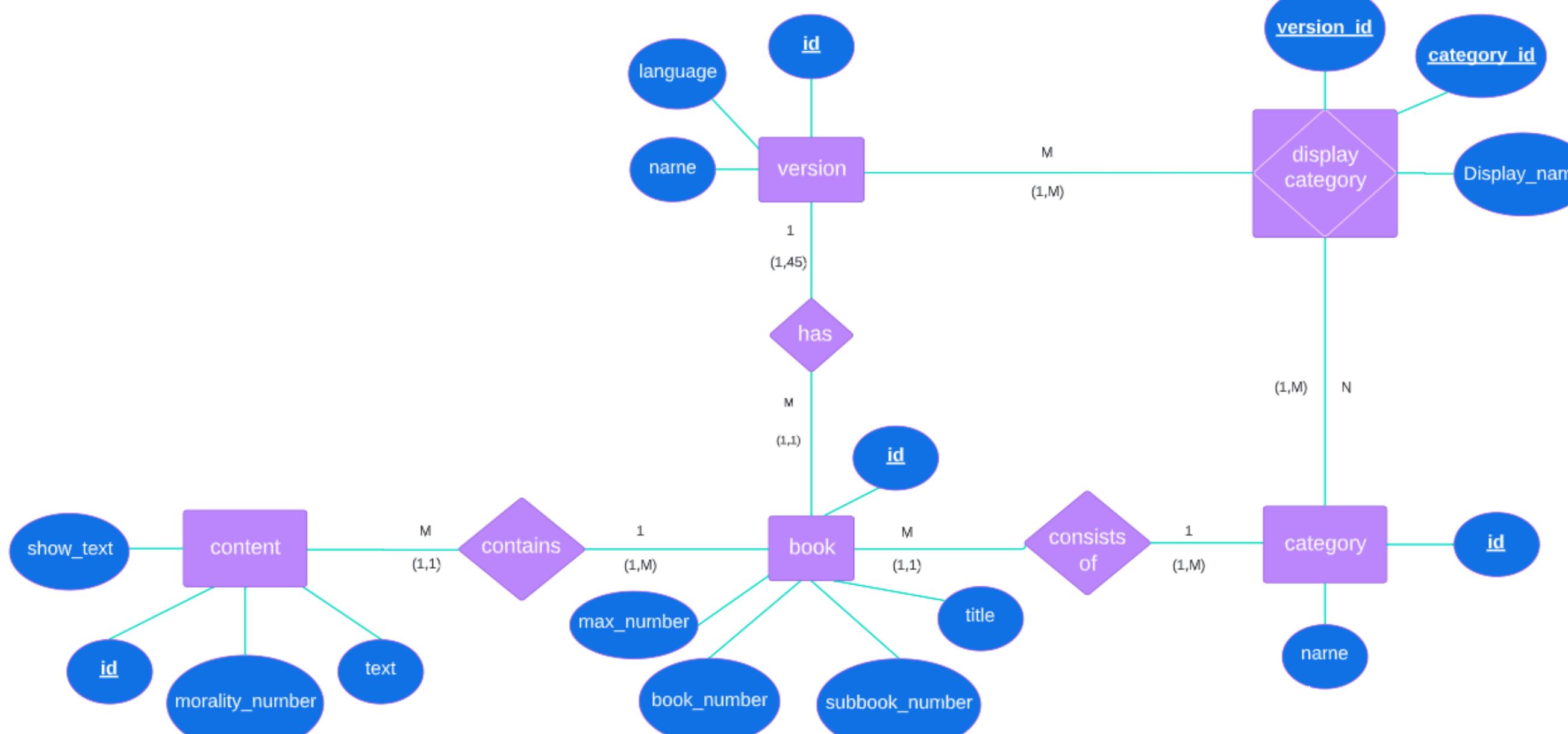
Sample data of JSON schema

book_information collection

```
[{  
    "version_id": 1,  
    "version": "ฉบับหลวง",  
    "category": [{  
        "id": "1",  
        "display_name": "พระวินัยปิฎก",  
        "book": [{  
            "id": 1,  
            "title": "มหาวิถี กาด ๑",  
            "max_number": 657  
        },  
        {  
            "id": 2,  
            "title": "มหาวิถี กาด ๒",  
            "max_number": 881  
        },  
        {  
            "id": 3,  
            "title": "กิกขุณีวิถี",  
            "max_number": 504  
        },  
        {  
            "id": 4,  
            "title": "มหาวิถี กาด ๔",  
            "max_number": 783  
        }]  
    }]  
}
```

NoSQL Schema

1 version = 1 collection



Sample data of JSON schema

text_v1 collection

```
{  
    "text_id": 1,  
    "version": "ฉบับหลวง",  
    "category": "พระวินัยปิฎก",  
    "title": "มหาวิกัค ภาค ๑",  
    "tripitaka_id": 1,  
    "morality_num": 1,  
    "show_text": "<center>พระวินัยปิฎก<br></center><center>เล่ม ๑<br></center><center>มหาวิกัค ปฐมภาค<br></center><center>ขอนอบน้อมแด่พระผู้มีพระภาคอรหันต์สัมมาสัมพุทธเจ้าพระองค์นั้น.<br></center><center>เวรัญชกัณฑ์<br></center><center>เรื่องเวรัญชพราหมณ<br></center> [๑] โดยสมัยนั้น พระผู้มีพระภาคพระพุทธเจ้าประทับอยู่ ณ  
 คงไม้สะเดาที่น้ำเพรียกษ์<br>สิงสถิต เขตเมืองเวรัญชา พร้อมด้วยกิกขุสังฆหมู่ใหญ่ประมาณ ๕๐๐ รูป  
 เวรัญชพราหมณ<br>ได้สัตบข่าวณัดแหน่ว่า ท่านผู้เจริญ พระสมณะโโคดมศากยบุตร ทรงผนวชาจากศากยตระกูล<br>ประทับอยู่ ณ  
 บริเวณต้นไม้สะเดาที่น้ำเพรียกษ์สิงสถิต เขตเมืองเวรัญชา พร้อมด้วยกิกขุสังฆ<br>หมู่ใหญ่ประมาณ ๕๐๐ รูป  
 ก็แลพระกิตติศัพท์อันงามของท่านพระโคดมพระองค์นั้น ขจรไปแล้ว<br>อย่างนี้ว่า พระผู้มีพระภาคองค์นั้น  
 ทรงเป็นพระอรหันต์แม้พระราหูนี่ ทรงตรัสรู้เองโดยชอบ<br>แม้พระราหูนี่ ทรงบรรลุวิชชาและจรมะแม้พระราหูนี่  
 เสด็จไปดีแม้พระราหูนี่",  
    "text": "พระวินัยปิฎก เล่ม ๑ มหาวิกัค ปฐมภาค ขอนอบน้อมแด่พระผู้มีพระภาคอรหันต์สัมมาสัมพุทธเจ้าพระองค์นั้น.  
 เวรัญชกัณฑ์ เรื่องเวรัญชพราหมณ [๑] โดยสมัยนั้น พระผู้มีพระภาคพระพุทธเจ้าประทับอยู่ ณ  
 คงไม้สะเดาที่น้ำเพรียกษ์ สิงสถิต เขตเมืองเวรัญชา พร้อมด้วยกิกขุสังฆหมู่ใหญ่ประมาณ ๕๐๐ รูป เวรัญชพราหมณ  
 ได้สัตบข่าวณัดแหน่ว่า ท่านผู้เจริญ พระสมณะโโคดมศากยบุตร ทรงผนวชาจากศากยตระกูล ประทับอยู่ ณ  
 บริเวณต้นไม้สะเดาที่น้ำเพรียกษ์สิงสถิต เขตเมืองเวรัญชา พร้อมด้วยกิกขุสังฆ หมู่ใหญ่ประมาณ ๕๐๐ รูป  
 ก็แลพระกิตติศัพท์อันงามของท่านพระโคดมพระองค์นั้น ขจรไปแล้ว อย่างนี้ว่า พระผู้มีพระภาคองค์นั้น  
 ทรงเป็นพระอรหันต์แม้พระราหูนี่ ทรงตรัสรู้เองโดยชอบ แม้พระราหูนี่ ทรงบรรลุวิชชาและจรมะแม้พระราหูนี่  
 เสด็จไปดีแม้พระราหูนี่"  
},
```

Sample data of JSON schema

text_v2 collection

```
{  
    "text_id": 1,  
    "version": "ฉบับบาลีอักษรไทย",  
    "category": "พระวินัยปิฎก",  
    "title": "วินย. มหาวิกุจโค (๑)",  
    "tripitaka_id": 1,  
    "morality_num": 1,  
    "show_text": "

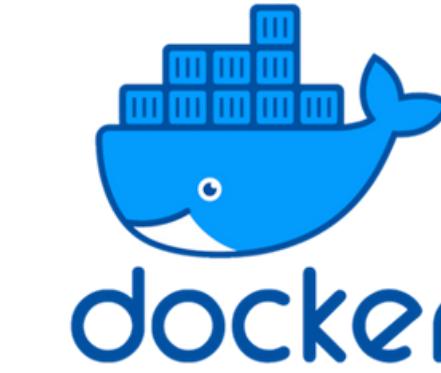
วินยปิฎก มหาวิกุจสุส<br>-----<br> นโน ตสส ภาโต อรหโต สมมาสมพุทธสุส  
ภาโค<br>ฯ<br> เวรญชกณฑ์<br> [๑] เตน สมเยน พุทธโธ ภาวา เวรญชาย วิหารด  
นเพรปุจิมनทุมเจ<br>มหดา กิกขุสูชน เ xen สหธี ปญจมตุเตหิ กิกขุสเตหิ ฯ อสุโสสิ<br>โข เวรญช  
พุราหมโน สมโน ชลุ ໂກ ໂຄตโน ສกุยปุตตโต ສກุยกุลja<br>ປພພชිໂຕ เวรญชาย วิหารด  
กิกขุสูชน<br>สหธี ปญจมตุเตหิ ต โข ปນ ກວນດໍ ໂຄດມໍ ເຂວ່ ກລຸຍາໂໂກ ກົດຕືສຫຼໂຫ<br>ອພກຸຄຸຄໂຕ ອິດີ ໂສ  
ภาวา ອຣໍ ສມມາສມພຸຖໂຣ ວິຊ້ຈາຈຮສນປນປນໂນ<br>ສຸຄໂຕ ໂລກວິຖ ອນດຸດໂຣ ປຸ່ຮິສທນມສາຮັດ ສັດຖາ ແວມນຸ້ສູສານ  
ພຸຖໂຣ<br>ກະວາຕີ ໂສ ອິມ ໂລກ ສເທວກ ສມາຮກ ສພຽມກຳ ສສຸສມພຸຣາຫຼົມ&lt;<br>ປ໌ ສເທວມນຸ້ສູສຳ ສຍ ອກິພູ້  
ສຈຸດິກດວາ ປເວເທດີ ໂສ ອົມມໍ ແເສດີ<br>ອາທິກລຸຍາ&gt; ມຊ່ເມກລຸຍາ&gt; ປົບໂຍສານກລຸຍາ&gt; ສາດຖິ  
ສພູຍຄຸນ&lt;<br>ເກວລປົບພຸ່ພັນ ປົບສຸທົ່ມ ພຣາມຈົບຍໍ ປກເສດີ ສາຊຸ ໂຂ ປນ ດຄຖຽປານ<br>ອຣທຳ ທສສນ ໂທດີ ແລ້ວ<br>"  
},


```

Implementation

Tools

PAGE 28

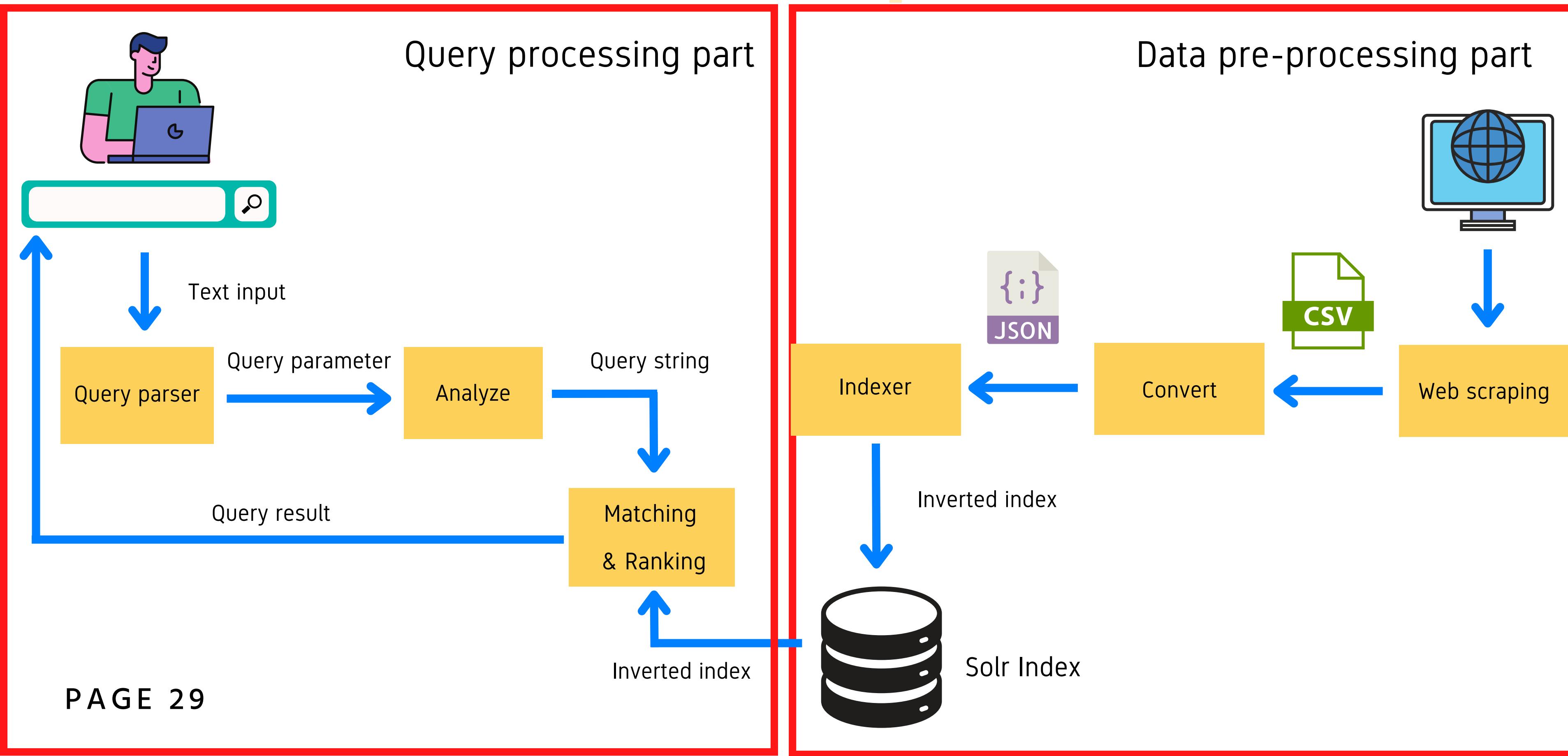


POSTMAN



DigitalOcean

Full-text search implementation



Data pre-processing part



Indexer process

1. Collect the documents to be indexed:

[១] ឯកសារលម្អិតដែលមានព័ត៌មានអនុញ្ញាតពីរបាយការណ៍ គឺជាបញ្ជីសម្រាប់ការបង្ហាញទូទៅ។

[២] ឯកសារលម្អិតដែលមានព័ត៌មានអនុញ្ញាតពីរបាយការណ៍ គឺជាបញ្ជីសម្រាប់ការបង្ហាញទូទៅ។

2. Tokenize the text, turning each document into a list of tokens:

១ ឯកសារលម្អិតដែលមានព័ត៌មានអនុញ្ញាតពីរបាយការណ៍ គឺជាបញ្ជីសម្រាប់ការបង្ហាញទូទៅ

២ ឯកសារលម្អិតដែលមានព័ត៌មានអនុញ្ញាតពីរបាយការណ៍ គឺជាបញ្ជីសម្រាប់ការបង្ហាញទូទៅ

3. Index the documents that each term occurs in by creating an inverted index, consisting of a dictionary and postings.

GENERATE POSTING

term	<docID, tf: <position of term in doc.>>
ອ	<1, 1:<1>>
ໄໂຍ	<1, 1:<2>>
ສນັບ	<1, 1:<3>>
ໜັນ	<1, 1:<4>>
ພຣະ	<1, 1:<5>>
ຖ້ວ	<1, 1:<6>>
ນີ້	<1, 1:<7>>
ພຣະ	<1, 1:<8>>
ກາຄ	<1, 1:<9>>
ພຣະຫຼອເຈົ້າ	<1, 1:<10>>
ປຣະທັນ	<1, 1:<11>>
ອຊູ	<1, 1:<12>>
ນ	<1, 1:<13>>
ຄວາງ	<1, 1:<14>>
ໄນ້	<1, 1:<15>>
ສະເດາ	<1, 1:<16>>
ເມ	<2, 1:<1>>
ໜັງຈາກ	<2, 1:<2>>
ໜັນ	<2, 1:<3>>
ເວັບ	<2, 1:<4>>
ນ	<2, 1:<5>>
ພຣະ	<2, 1:<6>>
ພຣະຫຼອເຈົ້າ	<2, 1:<10>>
ພຣະທັນ	<2, 1:<11>>
ກາຄ	<2, 1:<9>>
ໄລ້	<2, 1:<7>>
ໄປ	<2, 1:<8>>
ໃນ	<2, 1:<9>>
ປຣະທັນ	<1, 1:<11>>
ປຣະຫຼອຍ	<2, 1:<17>>
ໄປ	<2, 1:<8>>
ຖ້ວ	<1, 1:<6>>
ຖ່ານ	<2, 1:<10>>
ພຣະ	<1, 1:<5>>
ພຣະ	<1, 1:<8>>
ພຣະຫຼອເຈົ້າ	<1, 1:<10>>
ພຣະທັນ	<2, 1:<6>>
ກາຄ	<1, 1:<9>>
ນີ້	<1, 1:<7>>
ໄນ້	<1, 1:<15>>
ແລ້ວ	<2, 1:<14>>
ເວັບ	<2, 1:<4>>
ສນັບ	<1, 1:<3>>
ສະເດາ	<1, 1:<16>>
ສ້ານັກ	<2, 1:<11>>
ຄວິນ	<2, 1:<12>>
ຄື່ນ	<2, 1:<13>>
ແລ້ວ	<2, 1:<14>>
ໄລ້	<2, 1:<15>>
ຖ້ວ	<2, 1:<16>>
ປຣະຫຼອຍ	<2, 1:<17>>

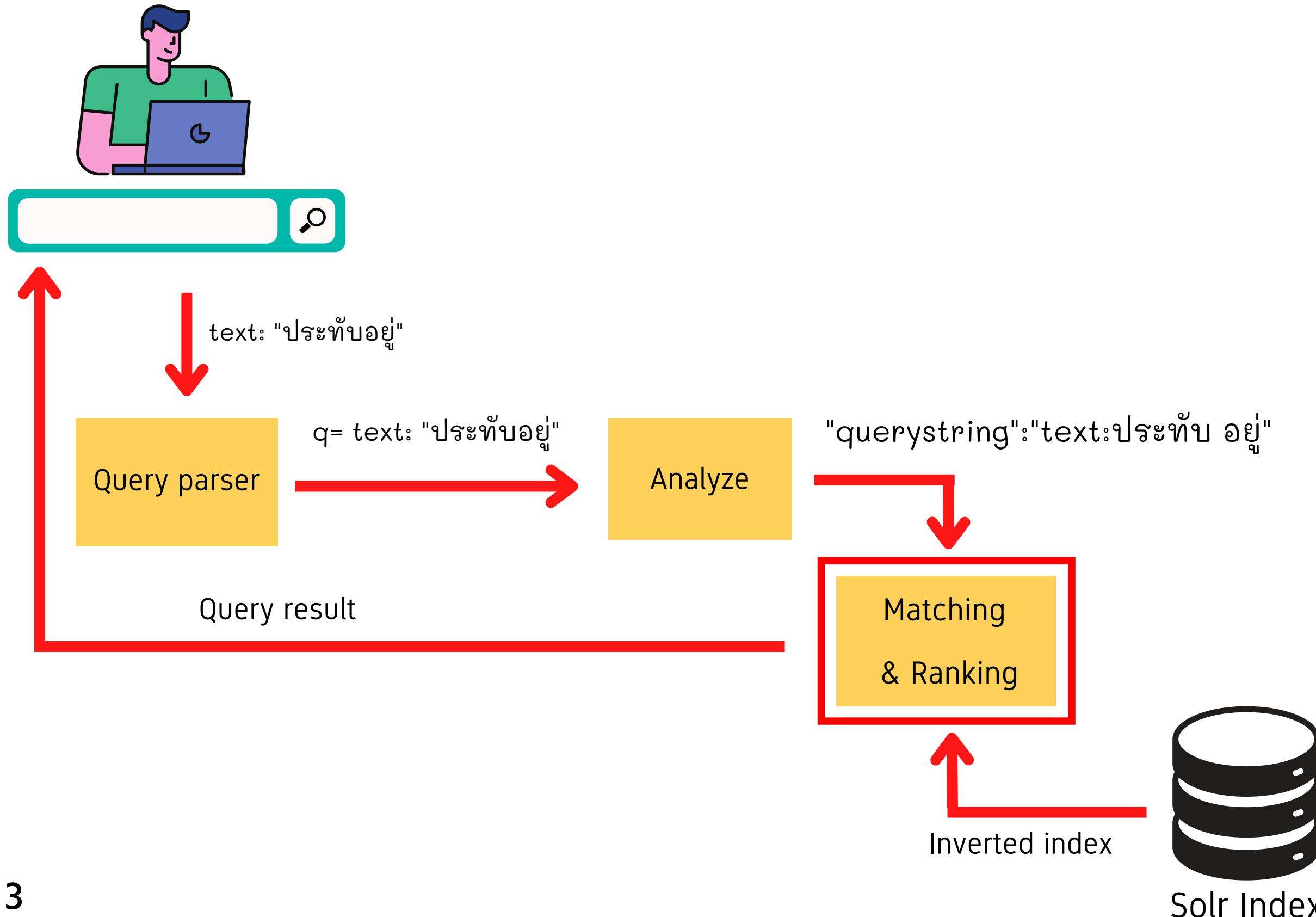
SORT POSTINGS

term	<docID, tf: <position of term in doc.>>
ຄວິນ	<2, 1:<12>>
ຄວາງ	<1, 1:<14>>
ນ	<2, 1:<5>>
ໄໂຍ	<2, 1:<5>>
ໄລ້	<2, 1:<7>>
ໄລ້	<2, 1:<15>>
ນ	<1, 1:<13>>
ຄື່ນ	<2, 1:<13>>
ຖ້ວ	<2, 1:<16>>
ນັນ	<1, 1:<4>>
ນັນ	<2, 1:<3>>
ໃນ	<2, 1:<9>>
ປຣະທັນ	<1, 1:<11>>
ປຣະຫຼອຍ	<2, 1:<17>>
ໄປ	<2, 1:<8>>
ຖ້ວ	<1, 1:<6>>
ຖ່ານ	<2, 1:<10>>
ພຣະ	<1, 1:<5>>
ພຣະ	<1, 1:<8>>
ພຣະຫຼອເຈົ້າ	<1, 1:<10>>
ພຣະທັນ	<2, 1:<6>>
ກາຄ	<1, 1:<9>>
ໄລ້	<1, 1:<7>>
ໄນ້	<1, 1:<15>>
ແລ້ວ	<2, 1:<14>>
ເວັບ	<2, 1:<4>>
ສນັບ	<1, 1:<3>>
ສະເດາ	<1, 1:<16>>
ສ້ານັກ	<2, 1:<11>>
ໜັງຈາກ	<2, 1:<2>>
ອຊູ	<1, 1:<12>>
ອ	<1, 1:<1>>
ເມ	<2, 1:<1>>

INVERTED INDEX (POSITIONAL INDEX)

term	df	<docID, tf: <position of term in doc.>>
ຄວິນ	1	<2, 1:<12>>
ຄວາງ	1	<1, 1:<14>>
ນ	1	<2, 1:<5>>
ໄໂຍ	1	<1, 1:<2>>
ໄລ້	1	<2, 2:<7, 15>>
ນ	1	<1, 1:<13>>
ຄື່ນ	1	<2, 1:<13>>
ຖ້ວ	1	<2, 1:<16>>
ນັນ	2	<1, 1:<4>> ; 2, 1:<3>>
ໃນ	1	<2, 1:<9>>
ປຣະທັນ	1	<1, 1:<11>>
ປຣະຫຼອຍ	1	<2, 1:<17>>
ໄປ	1	<2, 1:<8>>
ຖ້ວ	1	<1, 1:<6>>
ຖ່ານ	1	<2, 1:<10>>
ພຣະ	1	<1, 2:<5, 8>>
ພຣະຫຼອເຈົ້າ	1	<1, 1:<10>>
ພຣະທັນ	1	<2, 1:<6>>
ກາຄ	1	<1, 1:<9>>
ນີ້	1	<1, 1:<7>>
ໄນ້	1	<1, 1:<15>>
ແລ້ວ	1	<2, 1:<14>>
ເວັບ	1	<2, 1:<4>>
ສນັບ	1	<1, 1:<3>>
ສະເດາ	1	<1, 1:<16>>
ສ້ານັກ	1	<2, 1:<11>>
ໜັງຈາກ	1	<2, 1:<2>>
ອຊູ	1	<1, 1:<12>>
ອ	1	<1, 1:<1>>
ເມ	1	<2, 1:<1>>

Query processing part



MATCHING QUERY

QueryString: "ປະທັບ ອູ້"

1. Search for terms "ປະທັບ" and "ອູ້" in dictionary

ປະກັບ, 346

ອູ້, 2319

2. Access posting list of both term

ປະກັບ, 346:

<1, 1: <10>;
3, 3: <6, 20, 67>;
5, 2: <20, 54>;
10, 3: <15, 210>; ...>

ອູ້, 2319

<1, 1:<11>;
3, 3:<2,35,79>;
4, 5:<10, 42, 90, 111, 156>;
5, 4:<5, 15, 45, 140>;>

3. Match consecutive position of term "ປະກັບ" & "ອູ້" in the same documents.

ປະກັບ, 346:

<1, 1: <10>;
3, 3: <6, 20, 67>;
5, 2: <20, 54>;
10, 3: <15, 210>; ...>

Answer: Doc 1 <10>

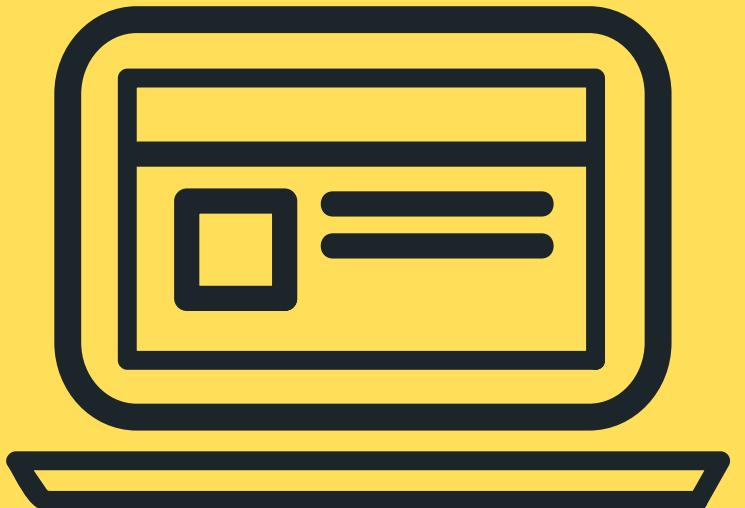
ອູ້, 2319

<1, 1:<11>;
3, 3:<2,35,79>;
4, 5:<10, 42, 90, 111, 156>;
5, 4:<5, 15, 45, 140>;>

MATCHING THE QUERY HAVE DONE.

WHAT'S ABOUT SOLR'S RANKING SCORE ?

RANKING SCORE BY BM25



WHAT IS BM25?

BM25 is a bag-of-words retrieval function that ranks a set of documents based on the query terms appearing in each document, regardless of their proximity within the document.

DISADVANTAGES OF TF-IDF

- Longer documents are given an unfair advantage over shorter ones
- Its inability in identifying more important and less important words for analysis.
- It will just consider words that are abundant in a corpus as the most statistically significant word.

BM25

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D)}{f(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)}$$

Where $f(q_i, D)$ = q_i 's term frequency in the document D .

$|D|$ = the length of the document D in words.

avgdl = the average document length in the text collection.

$b = 0.75$, length normalization parameter.

$k_1 = 1.2$, term saturation parameter.

$\text{IDF}(q_i)$ = the inverse document frequency of the i th query term.

$$\text{IDF}(q_i) = \ln \left(\frac{N - n(q_i) + 0.5}{n(q_i) + 0.5} + 1 \right)$$

N is the number of documents in the collection

$n(q_i)$ = the number of documents containing q_i .

EXAMPLE OF BM25 RANKING FUNCTION

QueryString = "เวร์ญชพราหมณ์", total documents = 3

D1 = [๑] หลังจากนั้น เวร์ญชพราหมณ์ได้ไปในพุทธสำนัก ครั้นถึงแล้วได้ทูลประธานร้อยก้าบ
พระผู้มีพระภาค ครั้นผ่านการทูลประธานร้อยพ่อให้เป็นที่บันเทิงเป็นที่ระลึกถึงกันไปแล้ว
จึงนั่ง ณ ที่ควรส่วนข้างหนึ่ง เวร์ญชพราหมณ์นั่ง ณ ที่ควรส่วนข้างหนึ่งแล้ว

D2 = [๒] เมื่อพระผู้มีพระภาคตรัสอย่างนี้แล้ว เวร์ญชพราหมณ์

D3 = [๓] ก็โดยสมัยนั้นแล เมืองเวร์ญชา มีกิจชาหารน้อย ประชาชนหาเลี้ยงชีพฝืดเคือง

|D|

59

14

18

so, avgdl = 30.33

$$\text{STEP 1: } \text{IDF}(q_i) = \ln \left(\frac{N - n(q_i) + 0.5}{n(q_i) + 0.5} + 1 \right)$$

"เวร์ญชพราหมณ์" occurred in 2 documents of 3 documents

$$= \ln \left(\frac{3 - 2 + 0.5}{2 + 0.5} + 1 \right)$$

$$= 0.47$$

STEP 2:

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)}$$

$$\text{score}(D1, Q) = 0.47 * \frac{2}{2 + 1.2 * (1 - 0.75 + 0.75 * 1.945)} = 0.232$$

$$\text{score}(D2, Q) = 0.47 * \frac{1}{1 + 1.2 * (1 - 0.75 + 0.75 * 0.461)} = 0.274$$

DEMO

TRIBUDDHA

WEB

APPLICATION



TEST SCENARIO FOR SEARCH SYSTEM

1. SINGLE WORD

General words

- พระ
- ปราสาท
- เลื่อมใส
- อาสนะ

Specific words

- โวมนิยา
- อัปปมัญญา
- สัปปจดยະ
- ຈາຣິກ

2. COMBINATION WORDS

General words

- ແມ່ໄປ
- ບາປ ບຸລູ
- ອາບຕີ ກົກຊູ
- ຮັງເກີຍຈ ສັບປຸງຮູ້

Specific words

- ອົມພລ້ວງສູງສຳກາ ກູ້ທັນຕະ
- ອຸກໂຕສຸຫະກັດ ລັກທາວະມານພ
- ປະມືຕະ ສັຈະກິກຕະ
- ອຸປລັມປາເປັກຂະ ວັນຕາຍິກອຮຽມ

3. CONCATENATE WORDS

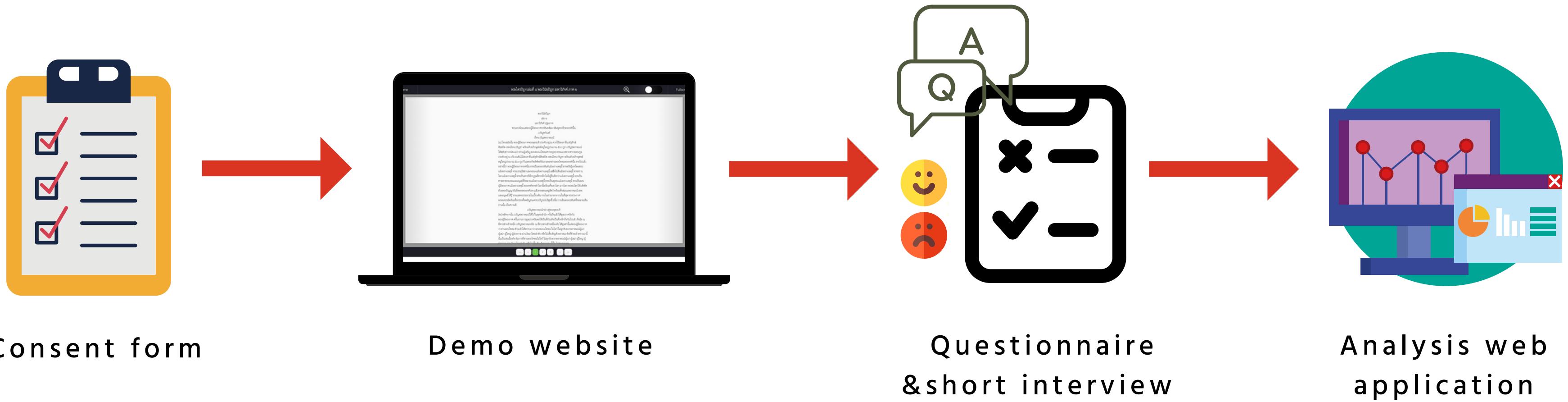
General words

- ກາບຖຸລວ່າ
- ດຸກຮົກສິກຊູ
- ບາຕຽງຈິວຮ
- ແມ່ນໍາຄອງຄາ

Specific words

- ພຣາມນົງກູ້ທັນຕະ
- ບຸ້າມຫາຍໍ້
- ເຂຕກຮູ່ງຮາຈຄູ້

Evaluation



Evaluation



Recruitment consist of
10 participants

5 participants

5 participants

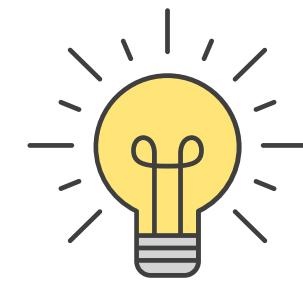
Experienced Monk

- Radbumrung temple 3 participants
- Thongnium temple 2 participants

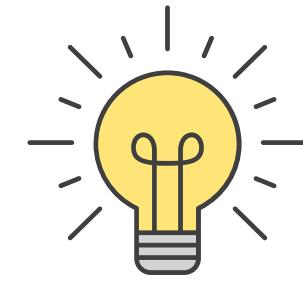
Unexperienced Monk

- Radbumrung temple 2 participants
- Thongnium temple 3 participants

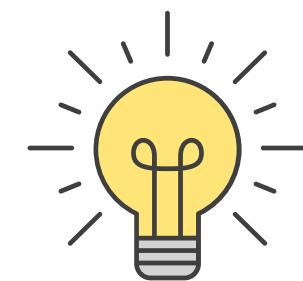
Evaluation



Functionalities



Overall satisfaction of TriBuddha website

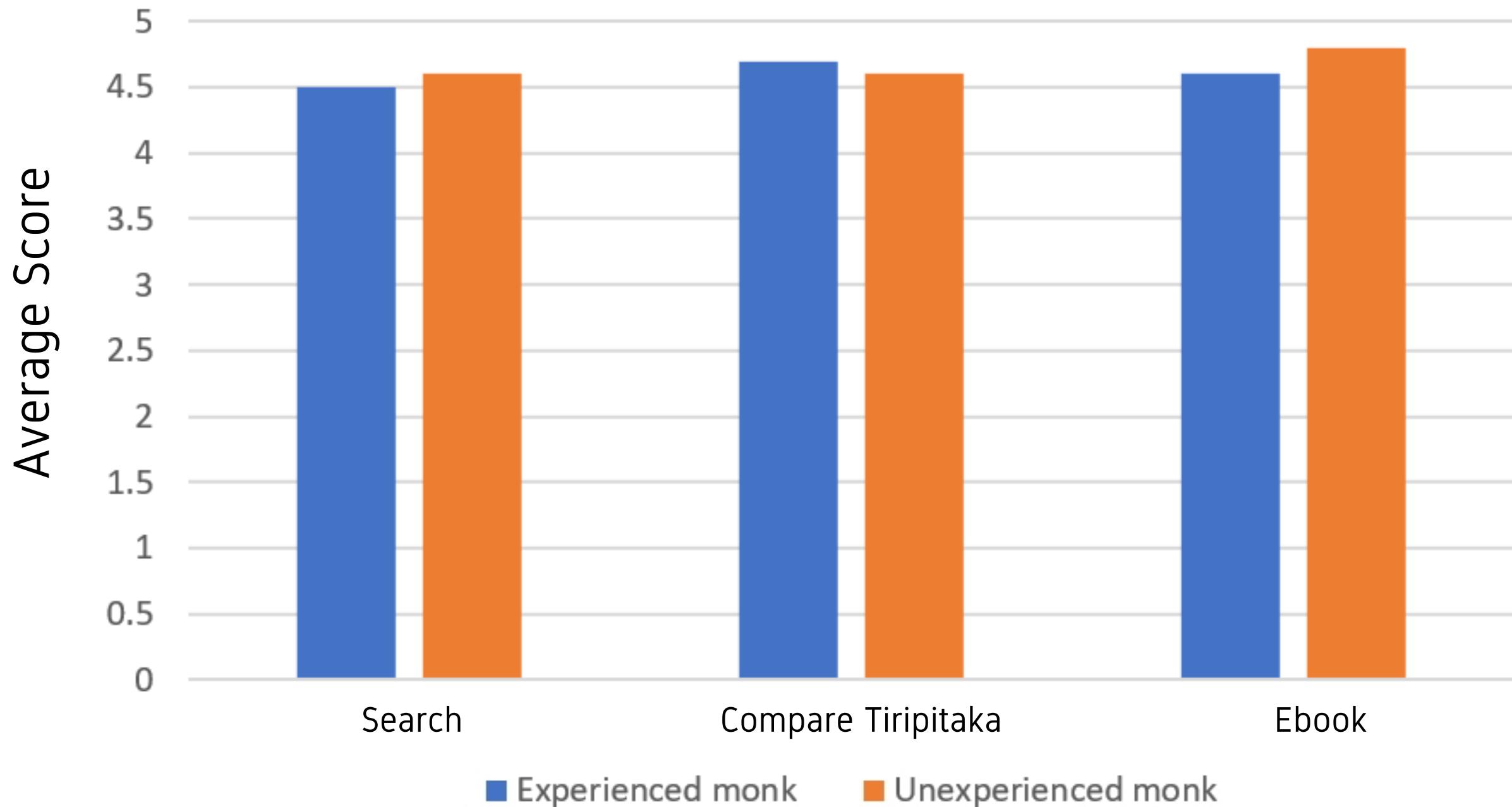


Satisfaction of TriBuddha website compared
to real paper Tripitaka book

Evaluation



Functionalities



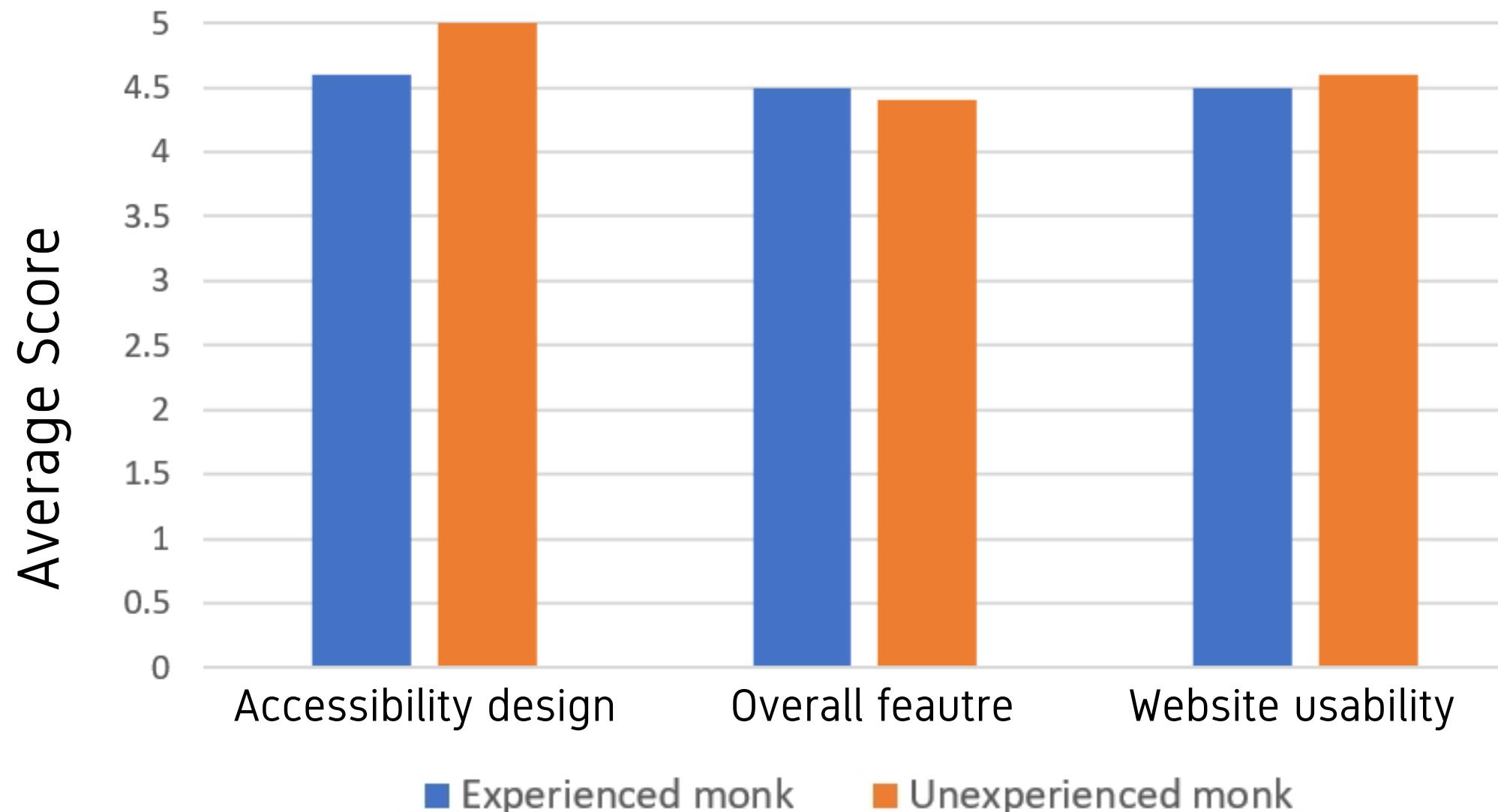
Summary

- all average score is equal or greater than 4.5
- Experienced monk gave a score to the compare Tripitaka at the highest score.
- Unexperienced monks most liked the Ebook feature.

Evaluation



Overall satisfaction of TriBuddha



Summary

- Both Experienced monk and unexperienced monk gave the accessibility design at the highest score
- Experience monk rated overall feature more than unexperience monk

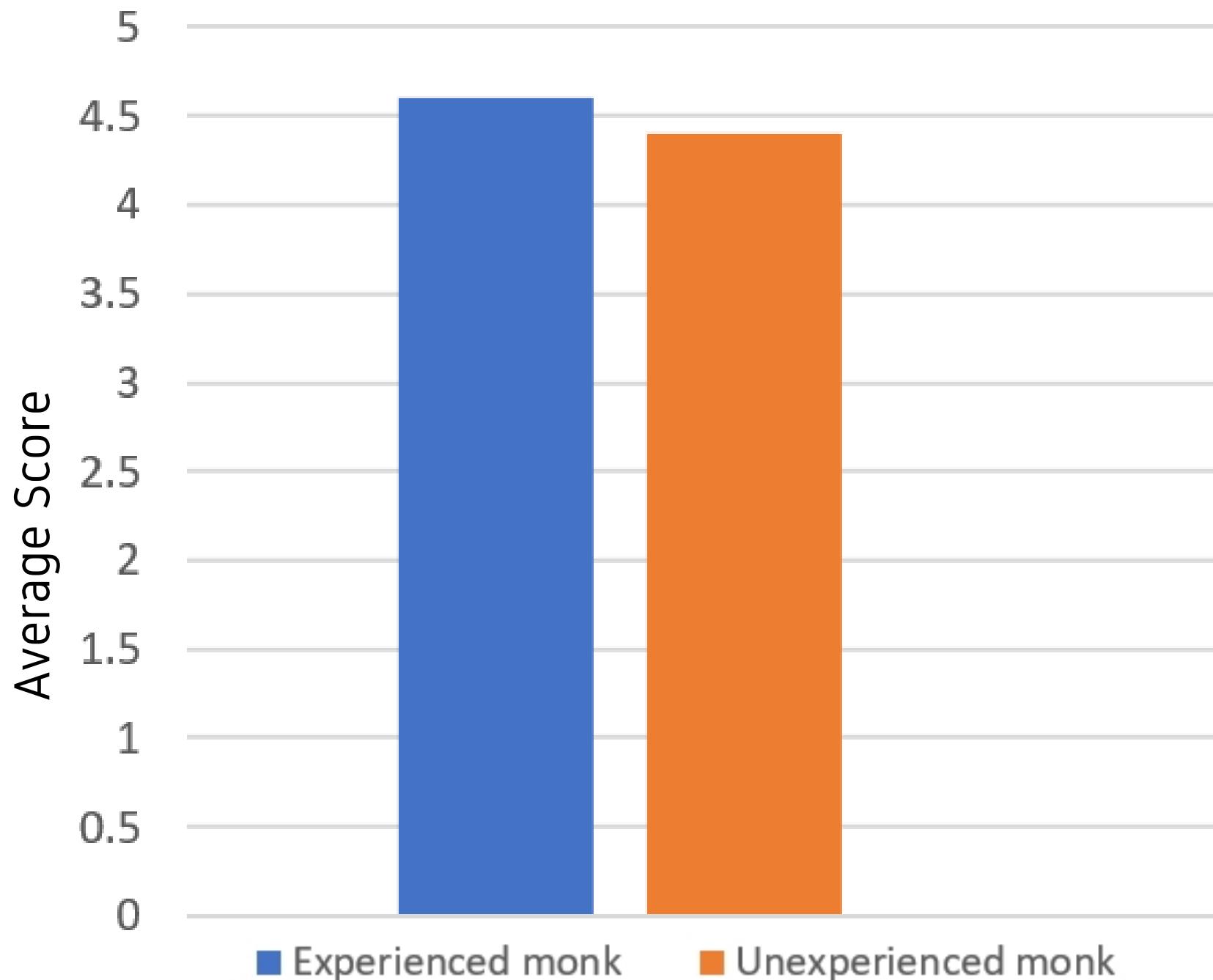
The website has clear content and has a day/night mode to protect his eye, and provide convenient and easy to find Tripitaka content.



Evaluation



Satisfaction of TriBuddha website compared to real paper Tripitaka book



Summary

- Experienced monk rated score more than unexperienced monk
- The averages of both groups were considered very good

very interested



Evaluation



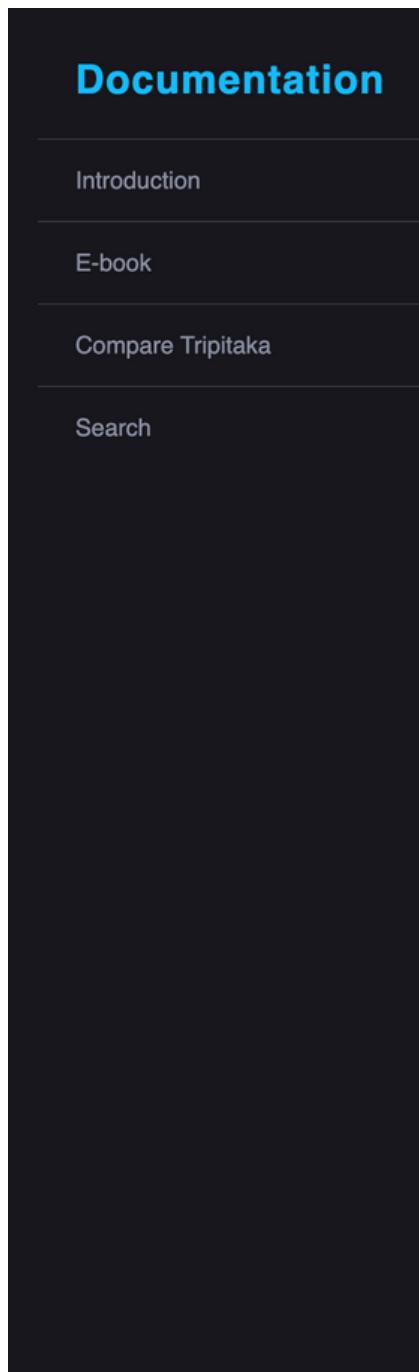
From interview, some of participant has suggestion and additional comment consists of;

- Add users manual in more detail
- Add Tripitaka in the version of Pali-Thai to Ebook feature

Evaluation



We have edited from the advice of the monk by add user manual in full version



Introduction

TriBuddha คือ เว็บแอพพลิเคชันที่รวมรวมพระไตรปิฎกทั้ง 45 เล่ม ซึ่งแบ่งออกเป็น 3 หมวดหมู่ คือ พะวินัยบัญญัค, พะสูตตันตบัญญัค และพะอภิธรรมบัญญัค โดย TriBuddha แบ่งออกเป็น 3 Features หนังสือไตรปิฎกทรอนิกส์ (E-book), เทียบเคียงพระไตรปิฎก (Compare Tripitaka) และค้นหาคำในพระไตรปิฎก (Search)

- หนังสือพระไตรปิฎกอิเล็กทรอนิกส์ (E-book) รวบรวมหนังสือพระไตรปิฎกเพื่อให้ผู้ใช้งานอ่าน
- เทียบเคียงพระไตรปิฎก (Compare Tripitaka) ผู้ใช้งานสามารถเทียบเคียงระหว่างบันทึกเพื่อเพิ่มเข้าใจเนื้อหาในพระไตรปิฎกอย่างถ่องแท้มากยิ่งขึ้น
- ค้นหาคำในพระไตรปิฎก (Search) ผู้ใช้งานสามารถค้นหาคำเลือกคุณที่ต้องการด้วยคำในพระไตรปิฎกและกรองเนื้อหาตามที่ต้องการ เพื่อให้ผู้ใช้งานสามารถเข้าถึงเนื้อหาได้อย่างรวดเร็ว แม่นยำ

หนังสือพระไตรปิฎกอิเล็กทรอนิกส์ (E-book)

วิธีการใช้งานหนังสือพระไตรปิฎกอิเล็กทรอนิกส์ (E-book) มีขั้นตอนดังนี้:

- ผู้ใช้งานสามารถเลือกเมนู HOME ที่แนบตัวบนน โดยหน้าเริ่มต้นเมื่อเข้า TriBuddha web application จะเป็นหน้า HOME เมื่อผู้ใช้งานเข้าสู่หน้า HOME
- หน้า HOME เป็นหน้าที่รวมรายการพระไตรปิฎก 45 เล่ม โดยจะแบ่งหมวดเป็น พะวินัยบัญญัค พะสูตตันตบัญญัค พะอภิธรรมบัญญัค ซึ่งสามารถ click แสดงเพิ่มเติม/แสดงน้อยลง เพื่อปรับจำนวนการแสดงผลตามความต้องการ
- เมื่อผู้ใช้งานเข้าสู่หน้าเนื้อหาพระไตรปิฎกแล้วแต่ละหน้าจะแสดงข้อความ เนื้อหาและปุ่มต่างๆ



- ปุ่มที่ 1: Home - เมื่อผู้ใช้งานต้องการที่จะออกจากหน้านี้หรือย้อนหนังสือเล่มอื่น สามารถกดปุ่ม HOME เพื่อกลับไปยังหน้า Home ที่รวมรายการพระไตรปิฎก 45 เล่ม
- ปุ่มที่ 2: วิจัยภาษา - ปุ่มนี้ใช้สำหรับเพิ่มขนาดตัวอักษรให้ใหญ่ขึ้น
- ปุ่มที่ 3: โหมดติดตาม - ปุมนี้ใช้สำหรับเปลี่ยนสีพื้นหลังเป็นสีดำและตัวอักษรเป็นสีขาว เพื่อถนอมสายตา
- ปุ่มที่ 4: Full screen - ปุมนี้ใช้สำหรับนำจอที่ประกอบด้วยๆ โดยขยายเนื้อหาพระไตรปิฎกออก

User satisfaction



Analysis web application



TriBuddha web application



Participants's Number of satisfaction
questionnaire

- The users are satisfied and give a very good rating score.

Benefits for users

- Users can reach Tripitaka content without a physical Tripitaka book
- To increase the channel for Tripitaka study to be easier to access
- Users reduce the time to find Tripitaka content by using TriBuddha function

Problems & Limitations

- We found that some keywords can not retrieve the Tripitaka contents although it exists.

ນັບຫລວງ ▾ ເລືອກທີ່ໜົດ ▾ ນໍາຍາຍ່ອງ

ພບທັງໝາດ 1 ຮາຍການ

ແສດທັງໝາດ	ພຣວິນຍິບູກ (0)	ພຣະສຸດຕິນທິບູກ (0)	ພຣະອົກົຽນມີບູກ (1)
1	ເລີນ 35 ພຣະອົກົຽນມີບູກ ວິດປະການ໌ ຂ້ອງ [ໃ] ຖຸກການ ເປົ້າໂນ ຖຸກ ຂອງສົດວັນນຸ່າ ທີ່ມໍາຫຼຸດເນີນ ນໍາເຫດຍື່ຍາຍາມ ນໍາເກົລືຍືຕ ນໍາ ຕ້າກນີ້ ໄນນໍາຍາຍ່ອງ ການ ຖັນວ່າການ ສມມືກັນວ່າການ ໄນນໍາປາກອນນາ ໄນນໍາກັກ ໄນນໍາຂອບໃຈ ໄດ້ແກ່ ຖຸກ ເຊີຍ ກົ່ນ ຮສ ໂພຍຸ້ພະ ປິເສດຖະກິນ ສົດວັນນຸ່າ ທີ່ມໍາຫຼຸດເນີນ ເປົ້າໂນ		

ແສດທັງໝາດ 1 - 1 ຮາຍການ ຈາກ 1 ຮາຍການ

ນັບຫລວງ ▾ ພຣະອົກົຽນມີບູກ ▾ ນໍາຍາຍ່ອງ

ພບທັງໝາດ 0 ຮາຍການ

ແສດທັງໝາດ	ພຣວິນຍິບູກ (0)	ພຣະສຸດຕິນທິບູກ (0)	ພຣະອົກົຽນມີບູກ (0)
-----------	----------------	--------------------	--------------------

ນັບຫລວງ ▾ ເລືອກທີ່ໜົດ ▾ ເຫດກຸງວາຈົດໆ

ພບທັງໝາດ 2 ຮາຍການ

ແສດທັງໝາດ	ພຣວິນຍິບູກ (2)	ພຣະສຸດຕິນທິບູກ (0)	ພຣະອົກົຽນມີບູກ (0)
1	ເລີນ 7 ພຣວິນຍິບູກ ຈຸລວັກ ກາຄ 2 ຂ້ອງ 324 ພຸກການມູນາອົກ ຕ້າກນີ້ ພຣະຜູ້ມີພຣະກາປະກະກັບອູ້ທີ່ເນື່ອຈາກພົວດີຕາມພຣະຫຼາຍາ - *ກົມຍີ ແລ້ວເສື້ອງທີ່ຈາກທາງກຸງວາຈົດໆ ເສື້ອງທີ່ຈາກໄປຕືອນສຳເນົາ ບັນຍາກຸງວາຈົດໆ ກວບວ່າ ພຣະທີ່ປະກະກັບອູ້ ດັ ພຸກການມູນາອົກ ຕ້າກນີ້ ພຣະຜູ້ມີພຣະກາປະກະກັບອູ້ທີ່ເນື່ອຈາກພົວດີຕາມພຣະຫຼາຍາ - *ກົມຍີ ແລ້ວເສື້ອງທີ່ຈາກທາງກຸງວາຈົດໆ ເສື້ອງທີ່ຈາກໄປຕືອນສຳເນົາ ບັນຍາກຸງວາຈົດໆ ກວບວ່າ ພຣະທີ່ປະກະກັບອູ້ ດັ ພຸກການມູນາວິຫາຍອືບເນີນສາມາທີ່ພຣະວາຫານແຫຍ່ອ ແກ່ກວະແຕ ເຫດກຸງວາຈົດໆທີ່ນັ້ນ ພຣະວິນຍິບູກ ເປົ້າໂນ		
2	ເລີນ 7 ພຣວິນຍິບູກ ຈຸລວັກ ກາຄ 2 ຂ້ອງ 358 ເງື່ອງພະເທົວທີ [ຕະເລີ] ດັກນັ້ນ ພຣະຜູ້ມີພຣະກາປະກະກັບອູ້ທີ່ເນື່ອໄກສົມພື້ ຕາມຫຼາຍາ ກົມຍີ ແລ້ວເສື້ອງທີ່ຈາກໄປທາງກຸງວາຈົດໆ ເສື້ອງທີ່ຈາກໄປໂຕຍສຳເນົາ ບັນຍາກຸງວາຈົດໆ ແລ້ວ ກວບວ່າ ພຣະທີ່ປະກະກັບອູ້ທີ່ ເວຸ້ງວິຫາຍ ອັນເປັນສາມາທີ່ພຣະວາຫານ ເຂົ້າແກ່ກວະແຕ ເຫດກຸງວາຈົດໆທີ່ນັ້ນ ພຣະວິນຍິບູກ ເປົ້າໂນ		

ແສດທັງໝາດ 1 - 2 ຮາຍການ ຈາກ 2 ຮາຍການ

ນັບຫລວງ ▾ ເລືອກທີ່ໜົດ ▾ ເຫດກຸງວາຈົດໆ

ພບທັງໝາດ 18 ຮາຍການ

ແສດທັງໝາດ	ພຣວິນຍິບູກ (2)	ພຣະສຸດຕິນທິບູກ (16)	ພຣະອົກົຽນມີບູກ (0)
1	ເລີນ 7 ພຣວິນຍິບູກ ຈຸລວັກ ກາຄ 2 ຂ້ອງ 619 ຜູ້ມີພຣະກາປ ຕ້າກນີ້ ໃຫ້ ທີ່ວ່າງເຈີນພົວດີຕາມພຣະຫຼາຍາ - *ກົມຍີ ແລ້ວເສື້ອງທີ່ຈາກທາງກຸງວາຈົດໆ ເສື້ອງທີ່ຈາກໄປຕືອນສຳເນົາ ບັນຍາກຸງວາຈົດໆ ກວບວ່າ ພຣະທີ່ປະກະກັບອູ້ ດັ ພຸກການມູນາອົກ ຕ້າກນີ້ ພຣະຜູ້ມີພຣະກາປະກະກັບອູ້ທີ່ເນື່ອຈາກພົວດີຕາມພຣະຫຼາຍາ - *ກົມຍີ ແລ້ວເສື້ອງທີ່ຈາກທາງກຸງວາຈົດໆ ເສື້ອງທີ່ຈາກໄປຕືອນສຳເນົາ ບັນຍາກຸງວາຈົດໆ ກວບວ່າ ພຣະທີ່ປະກະກັບອູ້ ດັ ພຸກການມູນາວິຫາຍອືບເນີນສາມາທີ່ພຣະວາຫານແຫຍ່ອ ແກ່ກວະແຕ ເຫດກຸງວາຈົດໆທີ່ນັ້ນ ພຣະວິນຍິບູກ ເປົ້າໂນ		
2	ເລີນ 7 ພຣວິນຍິບູກ ຈຸລວັກ ກາຄ 2 ຂ້ອງ 1 ພຣະວິນຍິບູກ ເປົ້າໂນ		

ແສດທັງໝາດ 1 - 2 ຮາຍການ ຈາກ 2 ຮາຍການ

Problems & Limitations

- We hypothesised that the ThaiTokenizer (default) can not tokenize terms of Tripitaka correctly.
- Change ThaiTokenizer to ICUTokenizer since has a group of researcher recommended.



Pairote Leelaphattarakij

to solr-user...@googlegroups.com

ถึงแม้ว่า Apache Solr เองนั้นรองรับการตัดคำภาษาไทยอยู่แล้วผ่านทาง component สามตัวคือ

- ThaiAnalyzer
- ThaiWordFilter
- ThaiTokenizer

แต่สำหรับผู้ใช้ที่ต้องการใช้ตัวตัดคำของ ICU ซึ่งมีข้อดีกว่าคือ

- รองรับการตัดคำหลายภาษาใน componentเดียว (เช่นภาษาอังกฤษ จีน ญี่ปุ่น)
- ตัดคำภาษาไทยได้แม่นยำกว่า



Pairote Leelaphattarakij

to solr-user...@googlegroups.com

ถ้าเปลี่ยนไปใช้ ICUTokenizer จะพอดัดคำได้นะครับ ดูในรูปด้านล่างนะครับ

text_th

"นาฬิกาคิดต์" > "นาฬิกาคิดต"

text_th_icu

"นาฬิกาคิดต์" > "นาฬิกา" "คิด" "ต"



rachu...@gmail.com

to solr-user...@googlegroups.com

สวัสดีครับ
ตอนนี้ผมใช้ Solr 4.8.1 วันนน Jetty ซึ่งใน Schema.xml ของผม config ไว้แบบนี้ครับ

```
<fieldType name="text_th" class="solr.TextField" positionIncrementGap="100">
  <analyzer>
    <tokenizer class="solr.ThaiTokenizerFactory"/>
    <filter class="solr.StopFilterFactory" ignoreCase="true" words="lang/stopwords_th.txt"/>
    <filter class="solr.WordDelimiterFilterFactory" generateWordParts="1" generateNumberParts="1" catenateWords="1" catenateNumbers="1" catenateAll="0"/>
    <filter class="solr.LowerCaseFilterFactory"/>
    <filter class="solr.ThaiWordFilterFactory"/>
    <filter class="solr.EnglishMinimalStemFilterFactory"/>
    <filter class="solr.RemoveDuplicatesTokenFilterFactory"/>
  </analyzer>
```

ผมมีปัญหาเรื่องของการ search คำว่า นาฬิกาคิดต ซึ่งตัว analyzer จะตัดคำให้ไม่ได้จริง ๆ แล้วมันควรแยกเป็น นาฬิกา + คิดต แต่มันกลับไม่สามารถตัดให้ได้เลย ส่งผลให้เวลาค้นคำว่า คิดต จะไม่มีผลค้าที่ซ่อน นาฬิกาคิดต ซึ่งมานะเลย ควรแก้ไขยังไงดีครับ

ราชันย



rachu...@gmail.com

to solr-user...@googlegroups.com

ขอบคุณมากเลยคํพคุณไฟ ໂຮງໝາ
ผมลองเอา text_th_icu ไปใช้สามารถตัดคำได้ดีกว่าจริง ๆ ด้วย

Problems and Limitations

ICUTokenizer

ICUT | น่า | ยกย่อง

ICUT | มี | ความ | รำคาญ

ICUT | เขต | กรุง | ราชคฤห์ | นั้น

ICUT | ทำ | ตาม | คำ | ของ | เจ้า | บ้าง

ICUT | บวช | บุตร | ที่ | มารดา | บิดา | มีได้

น่ายกย่อง

มีความรำคาญ

เขตกรุงราชคฤห์นั้น

ทำตามคำของเจ้าบ้าง

บวชบุตรที่มารดาบิดามีได้

ThaiTokenizer

TT | น่า | ก | ย่อง

TT | มี | วาม | รำคาญ

TT | เขต | กรุง | ราช | คฤห์นั้น

TT | ทำ | ตาม | คำขอ | ง | เจ้า | บ้าง

TT | บวช | บุตร | ที่มา | ร | ดา | บิดา | มีได้

Problems and Limitations

- We found that it can tokenize more accurate.
- However, it cannot tokenize some technical terms in Tripitaka.

ICUT | ດຣ | ແມ | ສັກ | ປື້ນ | ປກຄນ ດຣຮມລັ້ງນີ້ປກຄນ

ICUT | ວັດ | ຈມຣ | ຮຄ ວັດຈມຮຣດ

ICUT | ເສຍພ | ຮະກາ | ຍາ | ທາຣ ເສຍພຣະກາຍາທາຣ

- Another way is to add wordlist.
- Currently, Apache solar does not support of adding wordlist.

Future works

- Make TriBuddha responsive on Mobile phone.
- Add more versions of Tripitaka including ອຸປະກອດ and Pali-Thai on E-book.
- Provide more accessibility features including note, bookmark, and search history.

Conclusion

- Users can find original sources of the scriptures online by using the TriBuddha web application which has three core modules: search, comparison, and E-book.
- We conducted the test on monk 2 groups to evaluate the TriBuddha web application. The results were considered very satisfactory.
- It is an alternative for monks who have difficulty accessing the Tripitaka book and provide a more convenient search for the content of Tripitaka.



tribuddha.xyz



Thank you!

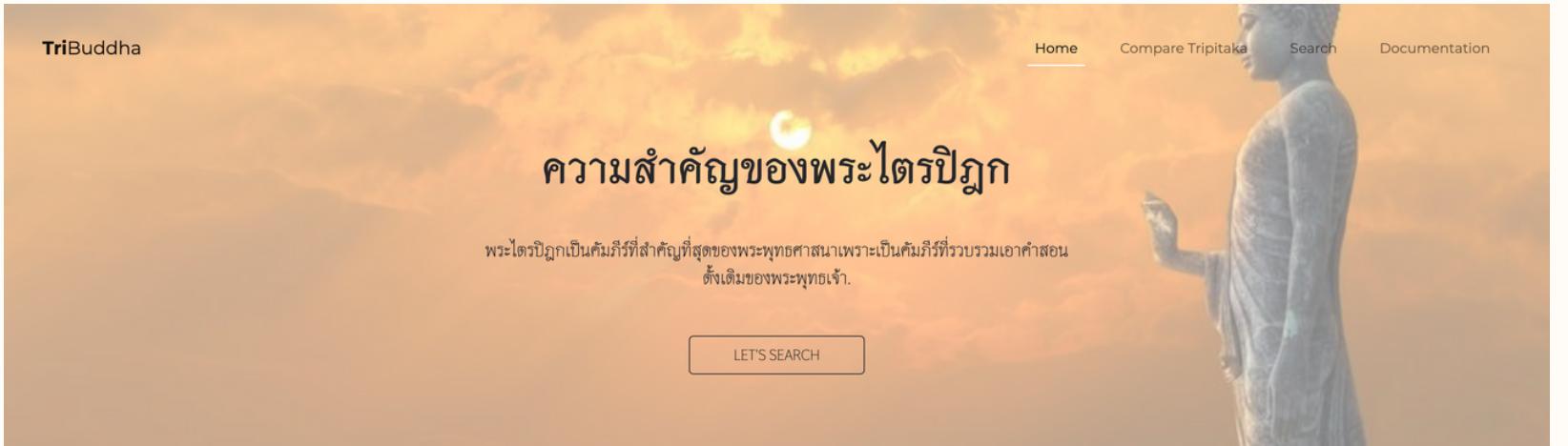




APPENDIX

E-Book

Homepage



1

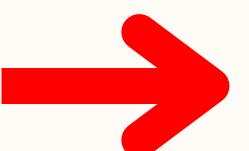
รวบรวมหนังสือพระไตรปิฎก 45 เล่ม

เลือกเพื่อการอ่าน

พระวินัยปิฎก



||||| แสดงเพิ่มเติม |||||



2

E-BOOK CONTENT

พระวินัยปิฎก
เล่ม 1
มหาวินัยปิฎก
ขอนอยบุรีและผู้มีพระภาคย์ที่นับถือในเชิงทางคันธ์
เว็บไซต์กัน
เรื่องราวในชีวิตประจำวัน

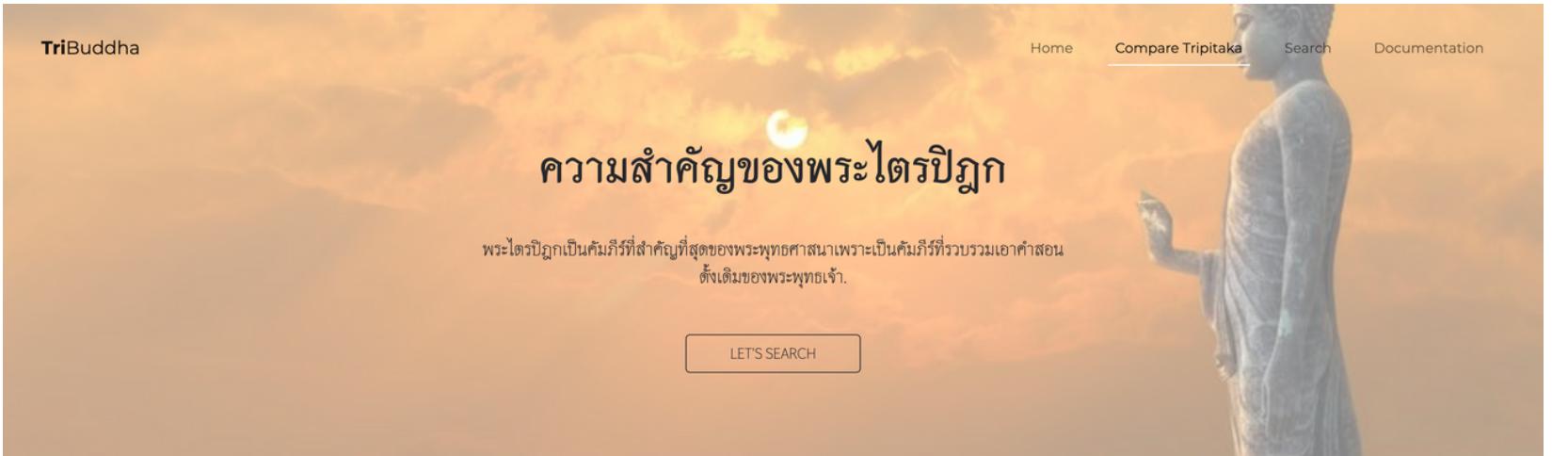
[๑] โดยสมัยนี้ พระผู้มีพระภาคย์ทรงทิ้งภารกิจอยุ กลางมืดมนาก็มีเชิงวิชาการที่มีชื่อว่า
สิงคิต เมืองเวรบุรา พิรุณตัวยักษ์ชั้นหนึ่งในญี่ปุ่นประมาณ ๕๐๐ ปี เว็บไซต์พากย์เสียง
ได้สัมภาษณ์แล้วว่า ท่านผู้เขียน พระสมบูรณ์โภคุมภากุล ทรงผ่านชาติภายนอกหลาย
ประทับอยุ บริเวณตัวเมืองเดิมที่นี่ ใช้ชื่อว่าสิงคิต เมืองเวรบุรา พิรุณตัวยักษ์ชั้นหนึ่ง
ที่ญี่ปุ่นประมาณ ๕๐๐ ปี กิมเบะกิตติศักดิ์ที่บ้านของท่านพระโคตมพะอย่างนั้น ชรับไปแล้ว
อย่างที่ว่า พระผู้มีพระภาคย์ที่นี่ ทรงเป็นพระอาทิตย์เมืองเวรบุรา ทรงครุฑ์อยู่โดยรอบ
เมืองเวรบุรา ทรงบรรลุวิชานะและธรรมะเพื่อความดูดี ทรงเป็น
โลกภัยเมืองเวรบุรา ทรงเป็นภารกิจไม่มีผู้อื่นอีกในเมืองเวรบุรา ทรงเป็น
ศาสต座ของทุกและบุษย์ทั้งหมดอย่างเมืองเวรบุรา ทรงเป็นพุทธะเมืองเวรบุรา ทรงเป็นพระ
ดูมีพระภาคย์เพื่อคนญี่ปุ่น พระองค์ทรงทำให้เมืองเวรบุรา เป็นเมืองที่มีความเจริญ
ด้วยพระบัญญาอีกซึ่งของพระคือเงา แล้วก็ทรงสอนหมู่สัตว์ พร้อมทั้งสอนมนุษย์ เทพ
และภูมิทั่วไป ทรงแสดงธรรมงานในเบื้องต้น ตามไปทั่วโลก จนมาในที่สุด ทรงประทาน
พระเมธราอย่างงามที่สุดของโลก ทรงยุบรวมและควบคุมบุปผาที่สุด ยังคงทรงเหลือเดิน
ป่านนี้ เป็นความที่

เว็บไซต์พากย์เสียง
[๒] หลังจากนั้น เว็บไซต์พากย์เสียงได้เป็นที่นิยมกันไปแล้ว ซึ่งมี ณ
พระผู้มีพระภาค ครั้นผ่านกาลเวลาเรื่อยๆ ให้เป็นที่นิยมกันไปแล้ว ซึ่งมี ณ
ที่ควรส่วนมากที่นี่ เว็บไซต์พากย์เสียงนั้น ที่ควรส่วนมากที่นี่ เว็บไซต์พากย์เสียงนั้น
ว่า ท่านพระโคตม ข้าพเจ้าได้ทราบมาว่า พระสมบูรณ์โภคุม ไม่ได้ไว้ไม่สุกในพากพากพากพาก
กันนั้น

<< < 1 2 3 > >>

COMPARE TRIPITAKA PAGE

Compare Tripitaka



ระบบเทียบเคียงภาษาในพระไตรปิฎก

เลือกอ่านเพื่อต้องการนี้เทียบเคียงและรับผลลัพธ์

1

2

SAMPLE RESULT

1. SINGLE WORD

Input= word : เลื่อมใส

The screenshot shows a search interface with a top bar containing dropdown menus for 'ฉบับหลวง' and 'เลือกทั้งหมด' (selected), and a search input field with the text 'เลื่อมใส'. Below the search bar is a link 'วิธีการใช้งาน'.

พนทั้งหมด 1259 รายการ			
แสดงทั้งหมด	พระวินัยปีฎก (427)	พระสุตตันตปีฎก (800)	พระอภิธรรมปีฎก (32)
1	เล่ม 8 พระวินัยปีฎก บริวาร ข้อ 1111 ว่าด้วยความเลื่อมใส [๑๗๑] ถ้าม่วาอย่างไร ชื่อว่าเลื่อมใสในสถานะควรเลื่อมใส? ตอบว่า กิจขุแสดงธรรมว่าธรรมชื่อว่ายอมเลื่อมใสในสถานะควรเลื่อมใส... กิจขุแสดงอาบติดชั่วทายนาว่าอาบติดชั่ว		
2	เล่ม 22 พระสุตตันตปีฎก อังคุตตรนิกาย ปัญจก-ฉักกนิبات ข้อ 218 ๔. อปานาทิกสูตร ๒ [๒๖๗] ดูกรกิจขุทั้งหลาย โทษ ๕ ประการนี้ มีอยู่ในบุคคลผู้ถึง พร้อมด้วยกิจกรรมเป็นต้นอันไม่น่าเลื่อมใส ๕ ประการเป็นไฉน คือ คนพาก ที่ยังไม่เลื่อมใสย่อมไม่เลื่อมใส ๑ คนพากที่เลื่อมใสแล้วบางพากย่อມคถาย ความเลื่อมใส ๑ ชื่อว่ายอมไม่กระทำตามคำสอนของพระศาสดา ๑ ประชุมชนชั้น		
3	เล่ม 36.2 พระอภิธรรมปีฎก ชาตุกถา-บุคคลบัญญัติปกรณ์ ข้อ 133 [๑๗๗] บุคคลผู้ถือรูปเป็นประมาณ เลื่อมใสในรูป เป็นไฉน บุคคลบางคนในโลกนี้ เห็นรูปที่สูง รูปที่ไม่อ้วน ไม่ผอม รูปที่มีอวัยวะ สมส่วน หรือรูปที่งามพร้อมไม่มีที่ติ ถือเอาเป็นประมาณในรูปนั้น แล้วจึงยังความเลื่อมใสให้เกิด นี้เรียกว่า บุคคลผู้ถือรูปเป็นประมาณ เลื่อมใสในรูป		

SAMPLE RESULT

2. COMBINATION WORDS

Input= word : ອາບຕີ ລົກໜູ້

SAMPLE RESULT

3. CONCATENATE WORDS

Input= word : ດາວໂຫຼດ

AVERAGE SCORES IN EACH SUB-QUESTION FROM 10 PARTICIPANTS

Website satisfaction information	Expereiced monk	Unexperience monk	Overall
1. Satisfaction of search for Tripitaka content by using keyword			
1.1 Fast and accurate access search system			
1.1.1 Single-word search	4.6	4.8	4.7
1.1.2 Combination search	4.8	4.8	4.8
1.1.3 Concatenate search	4.6	5	4.8
1.2 Data categorization	4.4	4.6	4.5
1.3 Comprehensive clearly content match with morality numbers	4.6	4.8	4.7
1.4 List of Tripitaka result completely	4.6	4.8	4.7
1.5 Easy-to-view list of search results	4.6	4.4	4.5
2. Satisfaction of search for Tripitaka content by using morality number			
2.1 Comprehensive clearly content match with morality numbers	5	5	5
2.2 Fast and accurate access result	4.6	4.8	4.7
2.3 Easy-to-view search result	4.6	4.2	4.4
3. Satisfaction of Tripitaka e-book form			
3.1 Comprehensive clearly content match with morality numbers	4.6	4.8	4.7
3.2 Data categorization	4.6	4.8	4.7
3.3 Fast and accurate access Tripitaka content	4.6	4.8	4.7
3.4 Suitablity features	4.6	4.8	4.7
4. Satisfaction of website accessibility			
4.1 User manual	4.6	5	4.8
4.2 Web layout and elements	4.6	5	4.8
4.3 Color contrast for better readability and appropriate font size	4.6	5	4.8
5. Satisfaction of TriBuddha website			
5.1 Ebook feature	4.6	4.4	4.5
5.2 Search feature	4.4	4.4	4.4
5.3 Compare Tripitaka feature	4.6	4.6	4.6
6. Website usability			
6.1 Ebook feature	4.6	4.8	4.7
6.2 Search feature	4.4	4.4	4.5
6.3 Compare Tripitaka feature	4.6	4.8	4.7
7. Satisfaction of TriBuddha website compared to Tripitaka book textbook form	4.6	4.6	4.6

Permission form

We requested permission to use
Tripitaka data from 84000.org

ขอความอนุเคราะห์ผู้พัฒนาเว็บไซต์ 84000 สนับสนุนการให้ข้อมูลพระไตรปิฎกเพื่อการศึกษา

OA

ORNPICHA ARUNSANGSIL

พฤ 17/3/2022 12:48



ถึง: DhammaPerfect@yahoo.com <dhammaperfect@yahoo.com>

สานะกิจ: SIRIPEN PONGPAICHET; CHATCHAPONG MINGMONGKHL; KANITTHA CHIEOCHANSILP

พวกเราทีมพัฒนาเว็บไซต์เพื่อการศึกษาพระไตรปิฎก นักศึกษาชั้นปีที่ 4 คณะเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยมหิดล โครงการของเรามีชื่อว่า "TriBuddha" เป็นเว็บไซต์เพื่อการเรียนรู้และเผยแพร่พระไตรปิฎก ด้วยการทำวิทยานิพนธ์เพื่อการจบการศึกษา จำเป็นต้องใช้ข้อมูลและเนื้อหาของพระไตรปิฎก ซึ่งทางเราเห็นว่าเว็บไซต์ 84000 มีข้อมูลที่ครบถ้วนและน่าเชื่อถือเป็นอย่างยิ่ง

ด้วยเหตุนี้ทางเรา นักศึกษาชั้นปีที่4 ผู้พัฒนาเว็บไซต์ จึงเรียนมาเพื่อขอความอนุเคราะห์การสนับสนุนการทำวิทยานิพนธ์ และขอความอนุญาตจากท่านเพื่อไม่เป็นการละเมิดลิขสิทธิ์ โดยทางผู้พัฒนาเว็บไซต์ยินดีจะชี้แจงเครดิตและแหล่งที่มาของข้อมูล

ผู้พัฒนาเว็บไซต์หวังเป็นอย่างยิ่งที่ท่านจะให้การพิจารณา และสนับสนุนการทำรายงานเว็บไซต์ดังกล่าว

ขอแสดงความนับถือ

นางสาวอรพิชา อรุณแสงศิลป์

นางสาวนิษฐา เชี่ยวชาญศิลป์

นายชัชพงศ์ มีงมลคล

มหาวิทยาลัยมหิดล

D

dhammaperfect <dhammaperfect@yahoo.com>

พฤ 17/3/2022 14:13



ถึง: ORNPICHA ARUNSANGSIL

ผู้จัดทำฯ ขออนุโมทนาและไม่ขัดข้องใดๆ เเลຍครับ

คำว่า อนุญาต ไม่มีสาระ

ขอแสดงความนับถือ

ผู้จัดทำ <https://84000.org/tipitaka/>

14:12 17/3/2565

...

ฉันไม่ขัดข้อง

โวด ใช้ได้เลย

ตกลง

!
ค่าแนะนำซื้อต้นมีประโยชน์หรือไม่ ใช่ ไม่

ตอบกลับ | ส่งต่อ

BACKUP

MYSQL VS APACHE SOLR

Comparison the functional features	MySQL	Apache Solr
Language analysis	x	/
Ranking algorithm	TF-IDF	BM25
Speed for large indices	limited	more scalable and faster
Faceted search	x	/
Hit highlighting	x	/

Tripitaka structure

- The book numbers, chapters, sub chapters are different.
 - The morality numbers begin with 1 every time according to book number

พระวินัยปีฎก
เล่ม ๑
มหาวิถังค์ ปฐมภาค
ขออนุบน้อมแด่พระผู้มีพระภาคอรหันต์สัมมาสัมพุทธเจ้าพระองค์นั้น.
เวรัญชกันท์
เรื่องเวรัญชพราหมณ์

[๑] โดยสมัยนั้น พระผู้มีพระภาคพระพุทธเจ้าประทับอยู่ ณ คงไม่สังเดาที่นี่เพรยักษ์ สิงสถิต เขตเมืองเวรัญชา พร้อมด้วยกิกขุสังฆ์หมู่ ใหญ่ประมาณ ๕๐๐ รูป เวรัญชพารามณ์ ได้สตันข่าวณดั่นแล้วว่า ท่านผู้เจริญ พระสมณะ โสดมศากยบุตร ทรงผนวชาจากศากยตระกูล ประทับอยู่ ณ บริเวณต้นไม้สังเดาที่นี่เพรยักษ์สิงสถิต เขตเมืองเวรัญชา พร้อมด้วยกิกขุสังฆ์ หมู่ ใหญ่ประมาณ ๕๐๐ รูป ก็แลพระกิตติศักพ์อันงามของท่านพระ โสดมพระองค์นั้น ขจรไปแล้ว อย่างนี้ว่า พระผู้มีพระภาคองค์นั้น ทรงเป็นพระอรหันต์แม้พระเหล่านี้ ทรงตรัสรู้เอง โดยชอบ แม้พระเหล่านี้ ทรงบรรลุวิชชาและจรณะแม้พระเหล่านี้ เสด็จไปดีแม้พระเหล่านี้ ทรงทราบ โลกแม้พระเหล่านี้ ทรงเป็นสารถฝึกบุรุษที่ควรฝึกไม่มีผู้อื่นยิ่งกว่าแม้พระเหล่านี้ ทรงเป็น ศาสดាខทางเทพและมนชย์ทั้งหลายแม้พระเหล่านี้ ทรงเป็นพಥะแม้พระเหล่านี้ ทรงเป็นพระ

พระวินัยปิกุล
เล่ม ๗

จุลวรรณ ภาค ๒

ขอนอนบันน้อมแด่พระผู้มีพระภาคอรหันตสัมมาสัมพุทธเจ้าพระองค์นั้น
ชุทหกวัตถุขันธะ

เรื่องพระฉัพพัคคី วัตรในการอาบน้ำ

Tripitaka structure

- Chapters, sub chapters are different.
- The morality numbers begin with 1 every time according to book number, and the book number is the same since book number relies on category. In other words, the left figure is Tripitaka book number 1 and the right figure is Tripitaka book number 8 which we stored each document as morality number. Therefore, if we use Tripitaka book number (1-45) & morality number, we can access every document.

พระวินัยปิกุล

เล่ม ๑

มหาวิภัค ปฐมภาค

ขอนอบน้อมแด่พระผู้มีพระภาคอรหันต์สัมมาสัมพุทธเจ้าพระองค์นั้น。

เวรัญชกัณฑ์

เรื่องเวรัญชพราหมณ์

[๑] โดยสมัยนั้น พระผู้มีพระภาคพระพุทธเจ้าประทับอยู่ ณ คงไม้สะเดาที่นเเพรญักษ์ สิงสถิต เขตเมืองเวรัญชา พร้อมด้วยภิกษุสงฆ์หมู่ ใหญ่ประมาณ ๕๐๐ รูป เวรัญชพราหมณ์ ได้สตับข่าวตนด้วยว่า ท่านผู้เจริญ พระสมณะ โโคดมศากยบุตร ทรงผนวชาจากศากยตระกูล ประทับอยู่ ณ บริเวณต้นไม้สะเดาที่นเเพรญักษ์สิงสถิต เขตเมืองเวรัญชา พร้อมด้วยภิกษุสงฆ์ หมู่ ใหญ่ประมาณ ๕๐๐ รูป ก็แลพระกิตติคัพท์อันงามของท่านพระ โโคดมพระองค์นั้น ขจรไปแล้ว อよ่างนี้ว่า พระผู้มีพระภาคองค์นั้น ทรงเป็นพระอรหันต์แม้เพระเหตุนี้ ทรงตรัสรู้เอง โดยชอบ แม้เพระเหตุนี้ ทรงบรรลุวิชชาและจรณะแม้เพระเหตุนี้ เสด็จไปด้วยแม้เพระเหตุนี้ ทรงทราบ โลกแม้เพระเหตุนี้ ทรงเป็นสารถีฝึกบูรุษที่ควรฝึกไม่มีผู้อื่นยิ่งกว่าแม้เพระเหตุนี้ ทรงเป็น

พระสุดตันตปิกุล

เล่ม ๑

ทีมนิกราย สีลขันธวรรค

ขอนอบน้อมแด่พระผู้มีพระภาคอรหันต์สัมมาสัมพุทธเจ้าพระองค์นั้น

๑. พระมหาลสุตร

เรื่องสุปปิยปริพากกันพระมหาทัตตามนพ

[๑] ข้าพเจ้าได้สตับนาอย่างนี้-

สมัยหนึ่ง พระผู้มีพระภาค เสด็จดำเนินทางไกรระหว่างกรุงราชคฤห์กับเมืองนาพันทา พร้อมด้วยภิกษุสงฆ์หมู่ ใหญ่ประมาณ ๕๐๐ รูป แม้สุปปิยปริพากก์ได้เดินทางไกรระหว่างกรุง ราชคฤห์กับเมืองนาพันทา พร้อมด้วยพระมหาทัตตามนพผู้อันเตวาลิก. ได้ยินว่า ในระหว่างทางนั้น. สุปปิยปริพาก กล่าวติพระพุทธเจ้า ติพระธรรม ติพระสังฆ์ โดยอเนกปริยา ล้วน พระมหาทัตตามนพผู้อันเตวาลิกของสุปปิยปริพาก กกล่าวชมพระพุทธเจ้า ชมพระธรรม ชม พระสังฆ์ โดยอเนกปริยา อาจารย์และอันเตวาลิกทั้งสองนั้น มีถ้อยคำเป็นข้าศึกแก่กันโดยตรง ฉะนี้ เดินตามพระผู้มีพระภาคและภิกษุสงฆ์ไปข้างหลังฯ ลำดับนั้น พระผู้มีพระภาคเสด็จ

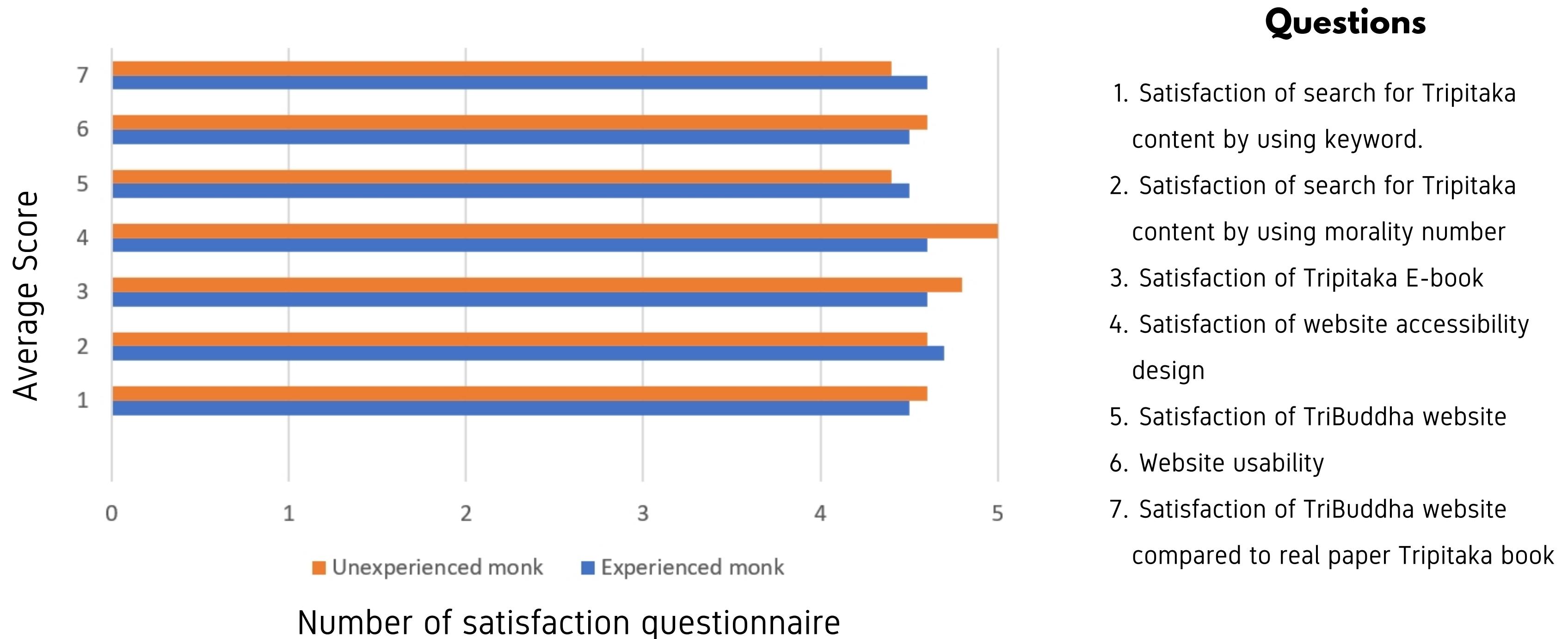
BENEFITS

Project Developers



- Experience in analyzing and verifying the data of Tripitaka.
- Cooperative skill.
- Experience of software design and development
- Skills in programming web application development

Evaluation



Evaluation



Experienced Monk + **Unexperienced Monk**

Summary

1. Satisfaction of search for Tripitaka content by using keyword.
2. Satisfaction of search for Tripitaka content by using morality number
3. Satisfaction of Tripitaka e-book
4. Satisfaction of website accessibility design
5. Satisfaction of TriBuddha website
6. Website usability
7. Satisfaction of TriBuddha website compared to real paper Tripitaka book



If k is set to 1, this would generate the sequence $1/2, 2/3, 3/4, 4/5, 5/6$ as TF increases 1, 2, 3, etc. Notice how this sequence grows fast in the beginning and then more slowly, approaching 1 in smaller and smaller increments. That's what we want. Now if we change k to 2, we'd get $1/3, 2/4, 3/5, 4/6$ which grows a little more slowly. Here's a graph of the formula $TF/(TF + k)$ for $k = 1, 2, 3, 4$:

