

# Uživatelská dokumentace – myRAG

## 1. Úvod

myRAG je aplikace pro sémantické vyhledávání a práci s dokumenty, která umožňuje rychle a efektivně získat odpovědi z vlastních textových zdrojů. Podporuje formáty PDF s textovou vrstvou, TXT a proprietární .alphadoc. Nahrané soubory se automaticky indexují a připraví k vyhledávání.

Aplikace využívá moderní přístup RAG (Retrieval-Augmented Generation), díky němuž kombinuje přesné vyhledávání relevantních pasáží textu s generováním odpovědí jazykovými modely. Výstupy jsou formulované přirozeným jazykem a opírají se o obsah nahraných dokumentů.

Minimální a doporučené hardwarové nároky pro běh aplikace myRAG:

| Parametr         | Minimální                             | Doporučené  |
|------------------|---------------------------------------|---|
| RAM              | 8GB                                   | 16GB+   |
| OS               | Windows 10 x64                        | Windows 11 x64, Ubuntu 24 amd64                                     |
| CPU              | 2.0 GHz čtyřjádrový procesor (x86-64) | 3.0 GHz osmijádrový procesor (x86-64)                               |
| Prostor na disku | 2GB                                   | 10GB+ (dle velikosti datasetů)                                      |
| GPU (volitelné)  | –                                     | NVIDIA GPU s CUDA, 6GB+ VRAM (urychlení výpočtů při práci s modely) |

Pozn.: Formát .alphadoc je proprietární souborový formát aplikace myRAG, který slouží k uložení celého projektu. V tomto jediném souboru jsou obsaženy všechny přidané dokumenty a jejich indexy pro vyhledávání. Uložení práce do .alphadoc souboru vám umožňuje kdykoliv projekt znovu otevřít a pokračovat v práci.



## 2. Co je RAG

RAG (Retrieval-Augmented Generation) propojuje vyhledávání informací (retrieval) s generováním textu (generation) pomocí jazykových modelů. Princip lze shrnout do dvou kroků:

- 1) Retrieval – Dotaz uživatele se porovná s databází dokumentů uložených ve vektorové podobě (embeddingy). Výsledkem je sada nejrelevantnějších úryvků (tzv. „chunků“).
- 2) Generation – Vyhledané úryvky jsou předány jazykovému modelu, který z nich sestaví odpověď v přirozeném jazyce.

Výhoda RAG: odpovědi nevznikají jen z obecných znalostí modelu, ale vycházejí přímo z vašich aktuálních a specifických zdrojů. To je zásadní pro odborné, proprietární i citlivé dokumenty.

## 3. Založení API klíčů

Pro poskytovatele vzdálených modelů (přes API) si založte API klíče:

- Anthropic Claude: <https://console.anthropic.com>
- OpenAI GPT: <https://platform.openai.com>
- Google Gemini: <https://aistudio.google.com>

Pozn.: Často je třeba zadat též zadat platební údaje a nastavit automatické dokupování tokenů pro nepřerušovaný provoz.

Pozn.: Ověřené modely, limitace a poznámky:

| Model                    | Poznámka                                       |
|--------------------------|--|
| gpt-4o                   | nejlepší pro jednodušší odpovědi               |
| claude-sonnet-4-20250514 | nejlepší pro komplexní odpovědi                |
| claude-opus-4-1-20250805 | nejlepší pro nejnáročnější odpovědi a problémy |
| gemini-2.5-flash         | nejlepší pro poměr rychlost odpovědi / kvalita |



|                |                          |
|----------------|--------------------------|
| gemini-2.5-pro | lepší model od googlu    |
| gpt-5*         | zatím nejsou podporovány |

## 4. Ollama (lokální inference)

### 4.1 Instalace

- 1) Stáhněte oficiální installer z <https://ollama.com> a spusťte ho.
- 2) Instalátor nainstaluje i službu/server Ollama, který běží na pozadí a může se spouštět automaticky po startu systému. Na Windows lze využít GUI, jinak lze ovládat přes terminál.

### 4.2 Rychlý test

V Terminálu nebo PowerShellu ověřte instalaci:

```
ollama --version
```

```
ollama pull gemma3:4b
```

```
ollama run gemma3:4b
```

Příkaz pull stáhne model, příkaz run otevře interaktivní sezení v terminálu.

### 4.3 Základní příkazy

ollama list – seznam stažených modelů

ollama pull X – stáhne model X

ollama run X – spustí model X (chat v terminálu)

ollama rm X – smaže model X

ollama cp A B – vytvoří kopii modelu A pod novým názvem B

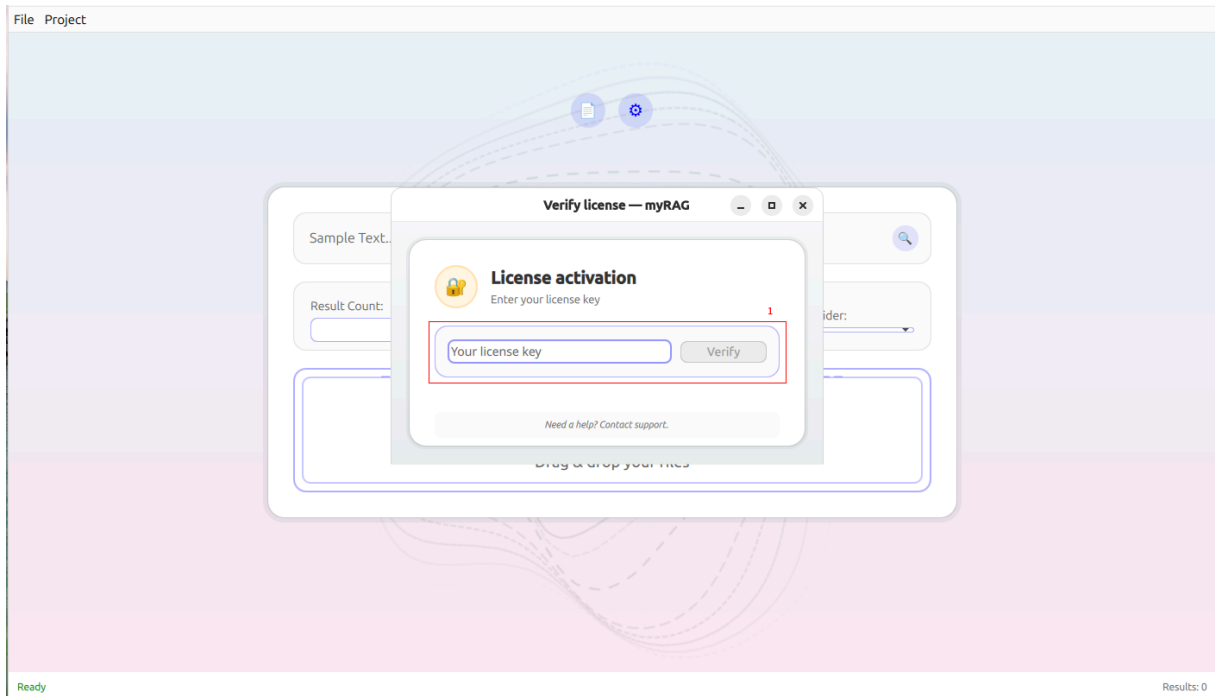
ollama serve – spustí REST API na <http://localhost:11434>

Pozn.: Na Windows je dostupná i oficiální GUI aplikace Ollama.

## 5. Práce s aplikací

### 5.1 První spuštění a aktivace licence

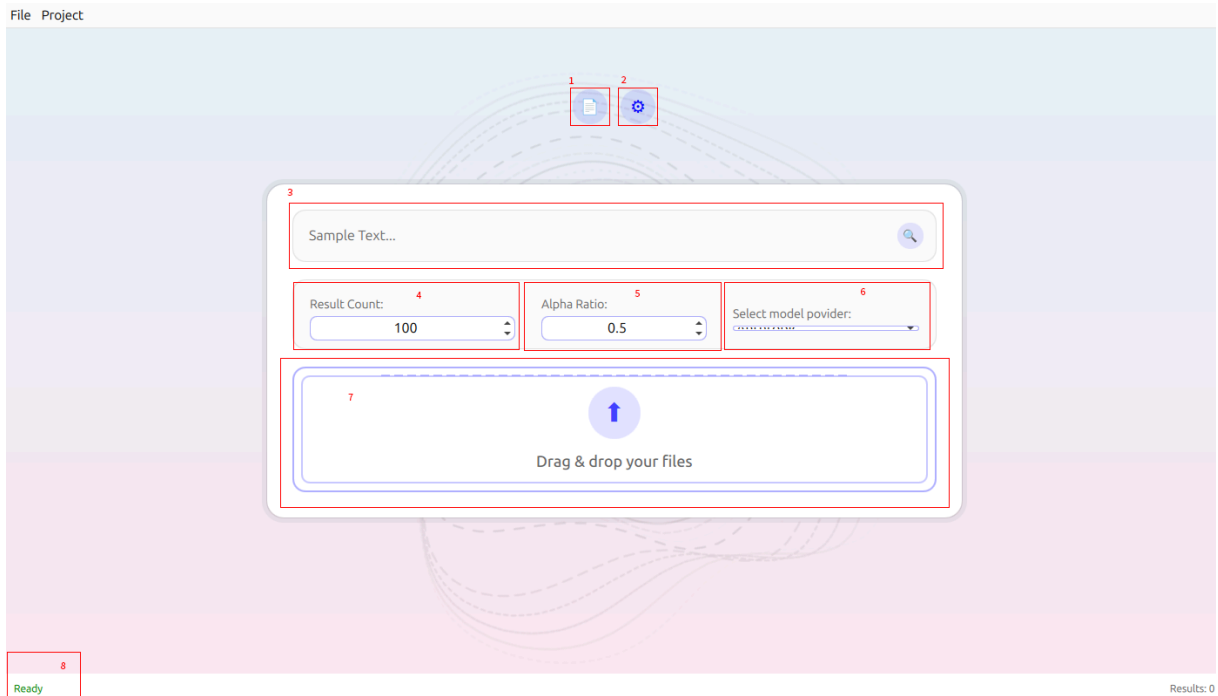
Při prvním spuštění budete vyzváni k zadání licenčního klíče. Po zadání správného klíče se aplikace odemkne.



Obrázek 1 – Okno pro aktivaci licence (pole „Your license key“).

## 5.2 Hlavní okno a jeho prvky

Toto je hlavní okno, ve kterém probíhá sémantické vyhledávání. Ovládací prvky (viz číslování na obrázku):



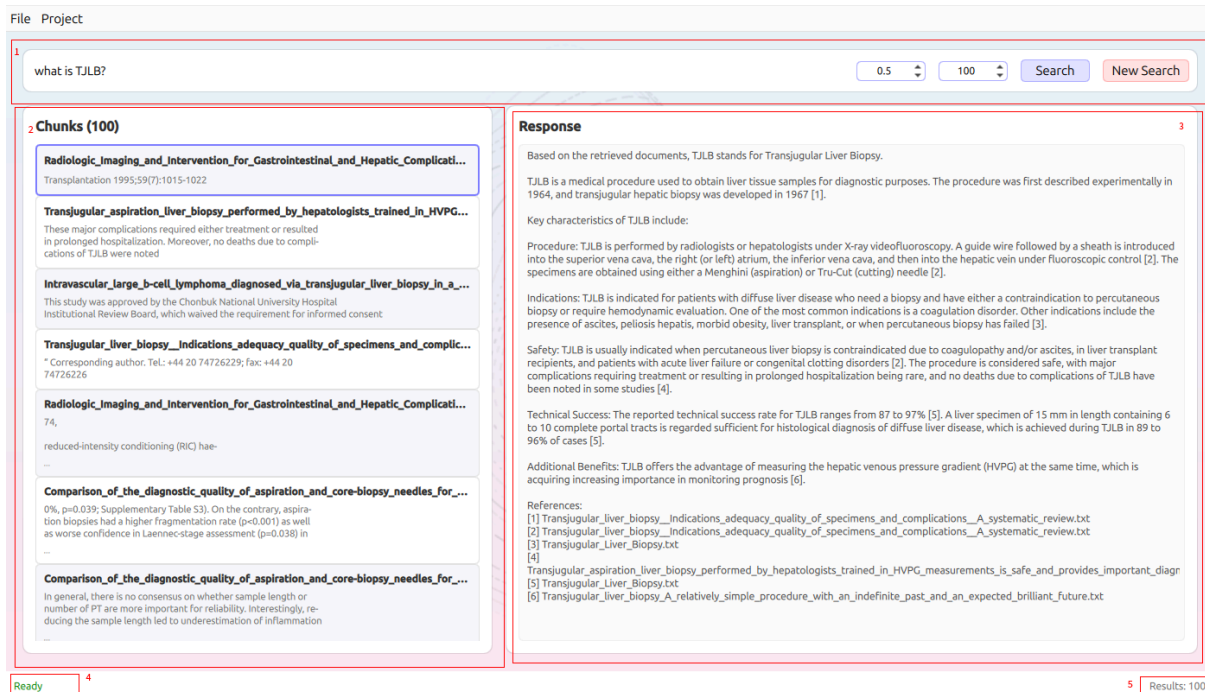
Obrázek 2 – Hlavní okno aplikace myRAG.

- 1) Ikona „Document Manager“ – otevře správce dokumentů.
- 2) Ikona „Settings“ – otevře nastavení aplikace.
- 3) Vyhledávací pole – napište dotaz a potvrďte Enter nebo kliknutím na lupu.
- 4) Result Count – počet navrácených chunků (pasáží) z vyhledávání.
- 5) Alpha Ratio – poměr fulltextového vs. sémantického vyhledávání. Hodnota 0 odpovídá čistě fulltextovému vyhledávání, zatímco hodnota 1 odpovídá čistě sémantickému vyhledávání. Hodnoty mezi 0 a 1 určují poměr mezi oběma typy.
- 6) Select model provider – volba poskytovatele/modelu (OpenAI, Anthropic, Gemini, Ollama...).
- 7) Drag & Drop area – drop-zóna pro soubory .pdf, .txt, .alphadoc.
- 8) Stavový pruh – indikace, zda aplikace právě pracuje (např. „Ready“, „Working...“).

Tip: Chcete-li otevřít uložený projekt, přetáhněte soubor .alphadoc do oblasti 7. Chcete-li přidat nové dokumenty, přetáhněte do oblasti 7 soubory .pdf nebo .txt.

### 5.3 Vyhledávání a zobrazení výsledků

Pro hledání stačí kliknout na ikonu lupy nebo stisknout Enter. Nejprve se zobrazí seznam nalezených chunků, následně (jakmile vybraný model vygeneruje výstup) se zobrazí formulovaná odpověď.

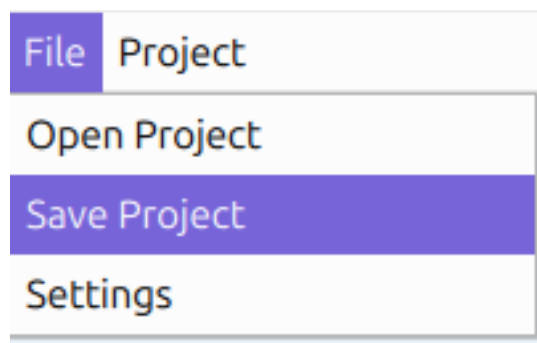


Obrázek 3 – Výsledky hledání: (1) dotaz a ovládání, (2) seznam chunků, (3) odpověď, (4) stav, (5) počet výsledků).

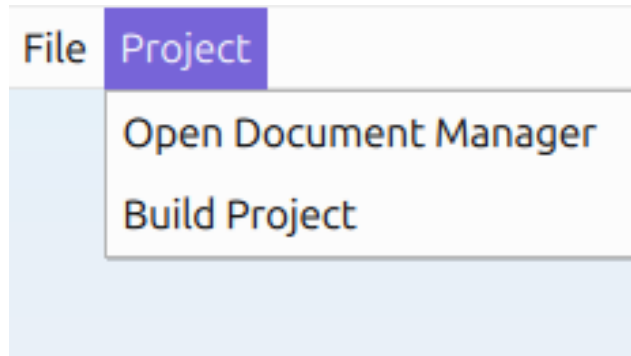
V záhlaví (1) lze zadat nový dotaz nebo zvolit „New Search“ a vrátit se na hlavní UI. Na stránce dále vidíte indikátor stavu aplikace (4) a počet vrácených chunků (5).

## 5.4 Menu aplikace

Záhlaví hlavního okna obsahuje menu s položkami „File“ a „Project“.



Obrázek 4 – Menu „File“ (Open Project, Save Project, Settings).



Obrázek 5 – Menu „Project“ (Open Document Manager, Build Project).

- Open Project / Save Project – práci s projektem (.alphadoc) otevřete/uložíte přes „File“.
- Settings – otevře nastavení aplikace.
- Open Document Manager – otevře správce dokumentů.
- Build Project – provede indexaci všech neindexovaných dokumentů projektu.

## 5.5 Nastavení

V nastavení zadáte parametry jednotlivých poskytovatelů, zvolíte aktuálního poskytovatele, spravujete licenci a upravujete systémový prompt pro formulaci odpovědí.

### Settings

**Actual AI provider**

.....

**Anthropic Claude**

Model:

API klíč:

**OpenAI GPT**

Model:

API klíč:

**Google Gemini**

Model:

API klíč:

**Ollama (local)**

Model:

URL:

**License**

Uživatel: Alpha Codes s.r.o.  
Detail: K. Mündla 513, 25230 Řevnice  
Platnost: 20.09.2026

[Change license](#)

**System prompt** To default

Given the query: {query}, and the following retrieved documents: {results}, answer the query using only the information provided in the documents. If the answer cannot be found in the provided results, respond with 'I don't know'.  
Base your answer strictly on the retrieved content and use scientific notation where appropriate. Cite each source using numbered references like [1], [2], etc., corresponding to a final list of unique document identifiers.  
At the end of your answer, include a numbered list of all used documents.  
Do not use any external knowledge or assumptions.  
Do not use markdown formatting in your response. Just plain text with new lines.

[Cancel](#) [Save](#)

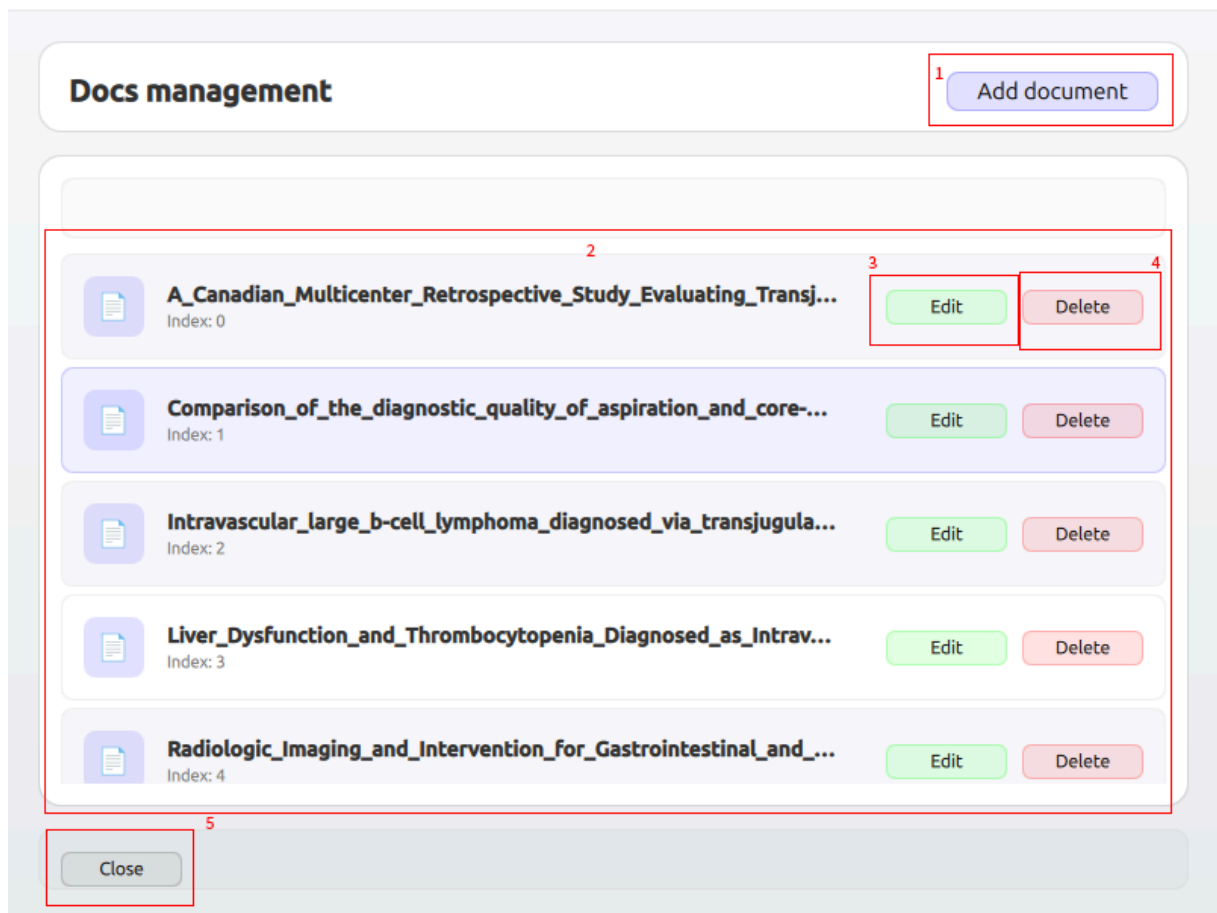
Obrázek 6 – Nastavení poskytovatelů (Anthropic, OpenAI, Google Gemini, Ollama local), Licence, System prompt.

Hlavní položky:

- Actual AI provider – aktivní poskytovatel/model pro vyhledávání a generování.
- Sekce poskytovatelů – „Model“ a „API klíč“ (u Ollama URL).
- License – informace o uživateli, detailech a platnosti licence; tlačítko „Change license“.
- System prompt – šablona, podle které je generována odpověď z nalezených dokumentů.

## 5.6 Správce dokumentů

Správce dokumentů slouží k přidávání, zobrazení, úpravě a mazání dokumentů v projektu.

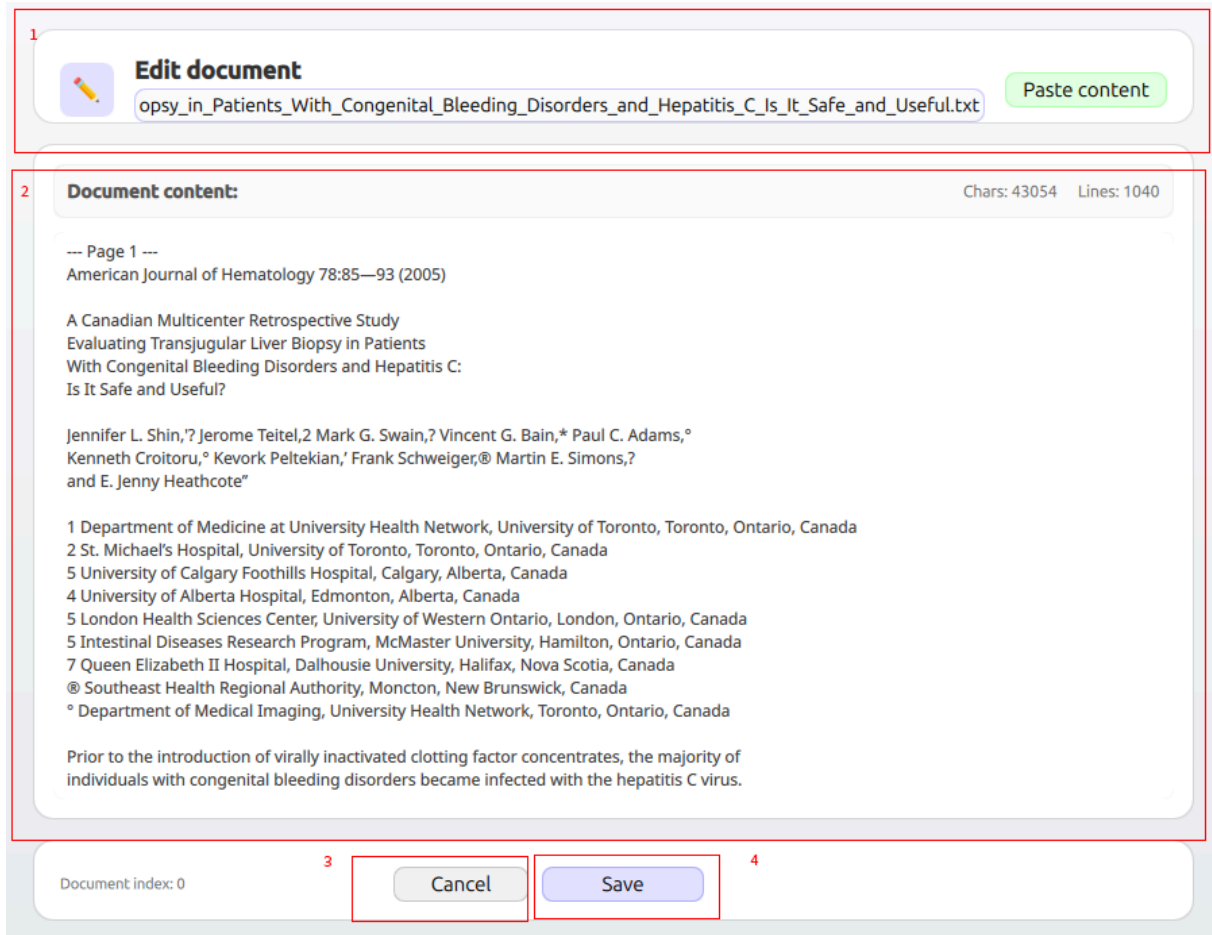


Obrázek 7 – Document Manager: (1) Add document, (2) seznam dokumentů, (3) Edit, (4) Delete, (5) Close.

Pozn.: Kliknutím na „Close“ se zavře správce dokumentů, ale projekt se automaticky neuloží – uložení provedte přes menu (viz výše).

## 5.7 Editor dokumentu

Editor umožňuje upravit název a obsah dokumentu a uložit změny.



Obrázek 8 – Editor dokumentu: (1) název souboru + tlačítko „Paste content“, (2) obsah dokumentu s počty znaků a řádků, (3) Cancel, (4) Save. Dole je uveden „Document index“.

Po uložení je třeba znovu „Build Project“, aby se změny projevíly ve vyhledávání (re-indexace).

## 6. Projekty (.alphadoc)

Projekt je uložen v jednom souboru s příponou .alphadoc. Lze jej otevřít přes „File → Open Project“ nebo přetažením .alphadoc do oblasti 7 na hlavní obrazovce. Uložení provedte „File → Save Project“.



Po přidání nebo úpravě dokumentů spusťte „Project → Build Project“, které provede indexaci nových/změněných dokumentů.

## 7. Tipy a doporučení

- Pokud nedostáváte dostatečně přesné odpovědi, upravte Alpha Ratio nebo zvýšte Result Count.
- Pro konzistentní odpovědi upravte System prompt – např. vyžadujte citace dokumentů a pracujte pouze s poskytnutými zdroji.
- Dokumenty rozdělte tématicky do projektů (.alphadoc) a pravidelně provádějte „Build Project“ po větších změnách.

## 8. FAQ a řešení problémů

- Aplikace nefunguje / padá vyhledávání: ověřte instalaci Ollama a dostupnost REST API na <http://localhost:11434>.
- Model negeneruje odpověď: zkontrolujte platnost API klíčů a to, že je vybráný provider v nastavení.
- Nepřichází relevantní výsledky: upravte Alpha Ratio a Result Count, případně proveďte „Build Project“.
- Uložení licencí / změn v nastavení: po úpravě klikněte na „Save“ v dialogu Nastavení.
- Změny dokumentu se neprojeví: po editaci vždy spusťte „Build Project“ (re-indexace).