

Experience with AI/ML Integration into Digital Libraries

...

Petr Žabička, Moravian Library in Brno
Jan Rychtář, Trinera

<https://bit.ly/AI-DL-Gdansk>



Moravská zemská knihovna v Brně - MZK.cz

- Legal deposit & research library
- [Knihovny.cz](https://www.knihovny.cz) portal of Czech libraries
- Mass digitization since 2012
 - 75 mil. pages in digital library
- Involved in many AI / DL projects
 - PERO OCR
 - Image identification and categorisation (Orbis Pictus)
 - Semantic segmentation and access (SemAnt)
 - AI creation of digitization metadata (Smart digiline)
 - Optical music recognition (OmniOMR)



Trinera

We help memory institutions efficiently manage data, automate processes, and make cultural heritage accessible to the wider public.

- Key partners: Moravian Library, National Library, and Library of the Academy of Sciences
- Our team consists of developers, librarians, UI/UX specialists, and AI/ML engineers
- Responsible for designing and developing the user interface of the Kramerius digital library
- We develop open source tools and applications for
 - Collection presentations
 - Metadata analysis
 - Library process automation using modern technologies, especially AI and ML
- AI-driven solutions include
 - Advanced search capabilities in digital libraries
 - Automated scan cropping in digitization
 - Extracting bibliographic data from title pages in cataloging
 - Enriching digitized documents

Introduction

- Digital libraries face new design and conceptual challenges
- Conceptual change from catalogue and fulltext search to semantic search and natural language queries
- Focus: user facing AI features

Kramerus

- Over 40 installations (CZ, SK, AT, GE)
- Open source digital library system
 - Java backend, Solr, IIP image server, Angular frontend
 - Jp2, pdf, epub, audio...
 - MODS, ALTO
 - Open API, IIIF image API, IIIF presentation API gateway
- Support for complex documents and hierarchical virtual collections
- Many types of access licenses incl. access for users with special needs

Czech digital library

- Kramerius-based common index of 15 largest Czech digital libraries
- Run by MZK in cooperation with nkp.cz and lib.cas.cz
- 292.000 books, 1.5 mil. periodical issues, maps, graphics, music, sound,...
- 23% public, 47% online after login (~ 1915-2004), 30% onsite access
- Aggregator for Europeana
- Full text indexed by Knihovny.cz Czech library portal

What Will Be Discussed

- Translation
- Text to speech
- Conversational interfaces, chat with a document
- Summarization
- Image description, image search
- Augmented retrieval and interpretation of search results

Translate page



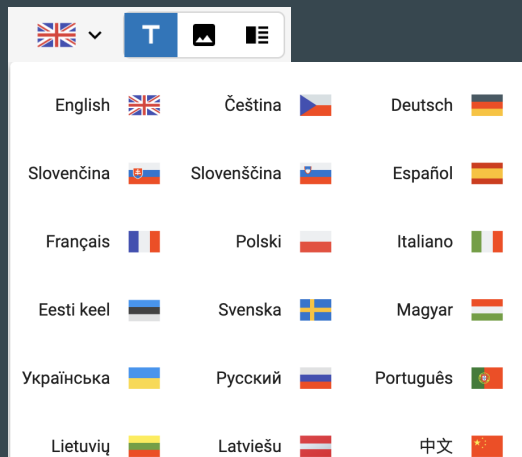
Translation

- Depends on OCR/HTR quality
 - Split words
 - Users might want to use different service or LLM
 - Some services do not support all languages
 - LLMs can work with transcription errors... up to a point
 - OCR might be AI corrected before translation
 - Translate page vs whole document
 - Incomplete sentences at page/column breaks
-

Translation

User must choose target language

- Not necessarily the same as UI language



Translation

Display of page:
scan/text/side by side

ken und drohten zu stürzen und ihn für immerdar zu verschütten.

Er fuhr sich mit der Hand über die Stirn. Seine Welt war grau und entfesselt worden, sein Herz war nimmer grün, und es wohnte wohl gar nicht mehr in seinem Leib, sondern lag dürr und vertrocknet in einem jener entsetzlichen Seelenmordenden Bücher.

Auf den fernen Wipfeln der Kindheit verglühte das Licht. Damals war er noch mit der Welt verbunden gewesen wie ein seliger Halm oder ein schuldloses Tier, damals hatte er mit Wind und Wassern gespielt und sich der Erde gefreut wie einer lachenden Mutter. Aber nun lag es wie ein graues Meer unüberschaubar zwischen ihm und jenen Zeiten.

Ihm war, als stünde er in einem klaren, kühlen Herbsttag, wo den Menschen erkennende Wehmut ergreift und ihm bang wird um sein dahinschwindendes Leben.

Er befühlte seine Hände, und ihn ekelte davor: diese tatenlosen, stillen Hände waren weich wie die einer behüteten Frau. Ihn ekelte auf einmal vor dem versponnenen Frieden am Schreibtisch. Aus dem engen Haus der Väter sehnte er sich dunkel ins Heimatlose, wünschte er sich wild und frei und obdachlos in ferne Erdteile, ohne Geld und Habe und Sicherheit, ausgehakt fremden Unwettern, auf Trieb und Schnelle und Kraft der Glieder angewiesen wie ein Tier. Ein neues,



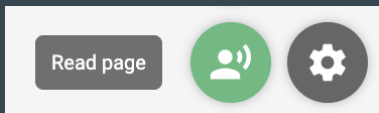
Ken and threatened to destroy him and bury him forever.

He ran his hand over his forehead. His world had become gray and lifeless; his heart was no longer green, and it probably no longer lived in his body, but lay dry and withered in one of those horrible, soul-murdering books.

On the distant treetops of childhood, the light faded. Back then, he had still been connected to the world like a blissful blade of grass or a guiltless animal; back then, he had played with wind and water and rejoiced in the earth like a laughing mother. But now, like a gray sea, it lay unfathomable between him and those times.

It was as if he were standing on a clear, cool autumn day, where a man's melancholy, recognizing it, seizes him and he becomes anxious about his fading life.

He felt delicate hands, and they disgusted him: these inactive, silent hands were soft like those of a sheltered witch. He was suddenly disgusted by the spun peace at the desk. From the narrow house of his ancestors, he longed darkly for homelessness, wishing to be wild, free, and homeless in distant parts of the world, without money, possessions, or security, exposed to strange storms, dependent on effort, speed, and the strength of his limbs like an animal. A new,



Text to Speech

- Read a document aloud
 - In what language (use translated text or the original?)
 - Rely on user interface language or translation language?
 - What model (voice) to use?
 - User selected languages are read without translation
 - Other source languages are translated to a primary language
-










Text to Speech

How to preprocess the text for reading

- Length of the text chunks
 - min / max
- Artificial spaces between the titles and the text
- Incomplete sentences at page/column breaks
- Automatic “page turning”

Settings / Text-to-speech

The listed languages will not be translated and the text will be read in the original language.
The remaining languages will be translated and read in the primary language.

Language	Voice	Play	
 English <input type="text"/>	▼ Fable F	 	
 Deutsch <input type="text"/>	▼ Thomas M	 	
 Čeština <input type="text"/> Primary	▼ Google F	 	
+ Add language			
<input type="radio"/> Fable F	<input type="radio"/> Alloy F	<input type="radio"/> Echo M	<input type="radio"/> Onyx M
<input type="radio"/> Nova F	<input type="radio"/> Shimmer F	<input type="radio"/> Biba F	<input type="radio"/> Honza M
<input type="radio"/> Rachel F	<input type="radio"/> Drew M	<input type="radio"/> Clyde M	<input type="radio"/> Paul M
<input type="radio"/> Domi F	<input type="radio"/> Dave M	<input type="radio"/> Fin M	<input type="radio"/> Sarah F
<input type="radio"/> Antoni M	<input type="radio"/> Thomas M	<input type="radio"/> Charlie M	<input type="radio"/> George M
<input type="radio"/> Emily F	<input type="radio"/> Elli F	<input type="radio"/> Callum M	<input type="radio"/> Patrick M
<input type="radio"/> Harry M	<input type="radio"/> Liam M	<input type="radio"/> Dorothy F	<input type="radio"/> Josh M
<input type="radio"/> Arnold M	<input type="radio"/> Charlotte F	<input type="radio"/> Alice F	<input type="radio"/> Matilda F
<input type="radio"/> James M	<input type="radio"/> Joseph M	<input type="radio"/> Jeremy M	<input type="radio"/> Michael M
<input type="radio"/> Ethan M	<input type="radio"/> Chris M	<input type="radio"/> Gigi F	<input type="radio"/> Freya F
<input type="radio"/> Brian M	<input type="radio"/> Grace F	<input type="radio"/> Daniel M	<input type="radio"/> Lily F
<input type="radio"/> Serena F	<input type="radio"/> Adam M	<input type="radio"/> Nicole F	<input type="radio"/> Bill M
<input type="radio"/> Jessie M	<input type="radio"/> Sam M	<input type="radio"/> Glinda F	<input type="radio"/> Giovanni M
<input type="radio"/> Mimi F	<input checked="" type="radio"/> Google F		



Ask a page / ask a document

- Experimental
 - Length of text vs. LLM choice (cost of query, time to answer)
 - LLM answers can be unpredictable, default prompt might help or bias the reply
 - Future work:
 - Chat-like interface
 - Chat with copyright protected document
-

Ask a page / ask a document

...er nun tag es wie ein graues, eckes, unabweisbares
zwischen ihm und jenen Zeiten.

Ihm war, als stünde er in einem klaren, kühlen
Herbsttag, wo den Menschen erkennende Wehmut er-
greift und ihm bang wird um sein dahinschwindendes
Leben.

Er befühlte seine Hände, und ihn ekelte davor:
diese tatenlosen, stillen Hände waren weich wie die
einer behüteten Frau. Ihn ekelte auf einmal vor dem

What does the narrator feel?

The narrator feels a deep sense of despair and disconnection from the world. He describes his existence as gray and soulless, indicating a loss of vitality and joy. There is a nostalgic longing for the innocence and connection he once felt during childhood, contrasting sharply with his current state of isolation and unease. He experiences a profound melancholy, as if he is standing on the brink of losing his life, and he feels repulsed by the complacency and inactivity around him. This leads to a yearning for freedom and a desire to escape to distant lands, embracing the rawness of life without the constraints of security or material possessions. Overall, the narrator's emotions are a mix of sadness, nostalgia, and a restless longing for something more authentic and alive.

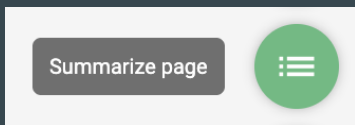


GPT 4o mini ▾

What does the narrator feel?



Close



Summaries

- User selected target language
 - Uses a translation to a target language to avoid LLM confusion
 - Predefined prompt
 - Currently at page or selection level
 - More logical at chapter/article level
-

Summaries

- Pre-generated AI summaries of full documents should improve user satisfaction
- In what language(s)?

The screenshot shows a user interface for generating AI summaries. A dropdown menu is open, listing several AI models:

- GPT 4o
- GPT 4o mini
- GPT 4o FT summary
- GPT 4o mini FT summary
- Claude 3 Haiku
- Claude 3 Sonnet
- Claude 3 Opus
- Claude 3.5 Haiku
- Claude 3.5 Sonnet
- Gemini 1.5 Pro
- Gemini 1.5 Flash
- Gemini 1.5 Flash 8B
- Gemini 2.0 Flash Experimental

Below the dropdown, a summary card is visible. It includes a language selector (United Kingdom flag) and a model selector (GPT 4o). The summary text is partially obscured but includes the following bullet points:

- Ken feels th...
- He experienc...
- Reflects on a...
- Feels a deep...
- Disgusted by...
- Desires to es...
- Longs for a l...

Additional text from the summary card includes: "of being buried forever.", "ected from his heart and past joys.", "innocence.", "without material possessions or security.", and "his own strength." A "Close" button is located at the bottom right of the interface.

TODO: Pre-generated summaries

- First question user might ask about a document
- Take long to create (> 10 s)
- Might help when semantic searching for a document / article
- Use cases:
 - Input for embedding generation
 - Become part of metadata
 - Help user quickly evaluate the document
- Each use case might require a different summary (in what language?)
- Named entity recognition and filter / facet generation might help as well

Image Description & Search - Orbis Pictus project

- Images and other non-textual objects on pages are identified and categorized
- Search by image similarity or textual description:
- Vector embeddings derived directly from the image
 - Search query must be processed by the same model
 - Might bring to the front unexpected/unwanted pictures, without context
- LLM generated textual description
 - Can be used by both full text and semantic search engines
 - Depends on use case (keywords vs short annotation vs detailed description)
 - These might be indexed into the full text index

- Image Search
- About the demo

skeleton riding on a fly

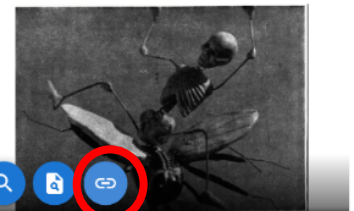
Query language
czech

- Libraries
- KNAV
 - NKP
 - MLP
 - VSE
 - UZEI
 - CUNI
 - SVKHK
 - SVKUL
 - NFA
 - KVKLI
 - NTK
 - SVKOS
 - NACR
 - D
 - KKVHB
 - ZMP
 - NLK
 - MKCT
 - VUGTK
 - NPMK
 - NULK
 - VSUP
 - CUNI-LF1
 - KKP
 - CUNI-FSV
 - MJH

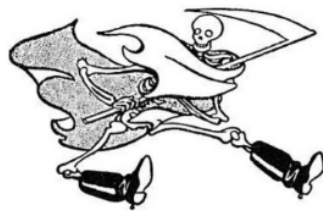
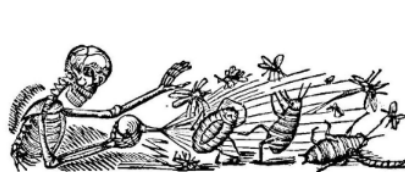
Retrieved images

Found 128 results in 2,357s

translation: 0,114s embedding: 0,022s search: 2,221s finalization: 0,00s



Open in Kramerius



Search in document

Page 7 / 44



You do not have access to this document now

The document is available under the licenses: [Study \(1\)](#)

To display it:

(1)

The document is not publicly accessible

The document you requested may not be freely viewed in accordance with [Czech Copyright Law \(No. 121/2000 Coll.\)](#).

The Copyright Act allows online access to works 70 years after the death of all authors (including translator, illustrator, etc.) and 50 years after the publication of the work, up to 1st January of the following year. Monographs, newspapers and magazines, for which the date of the authors' death is unknown, are made available 110 years after publication.

If you need more information or you think that any of our documents is not publicly accessible while it should be, contact us: knihovna@mjp.cz

Program D

Divadelní časopis / Umělecký měsíčník

ISSN 2336-6362

[Open periodical](#)

Volume

Year of publication **1946-1947**

Volume **10**

[Open periodical volume](#)

Issue

Date of publication **16.6.1947**

Issue **10**

[Previous issue 9 \(14.5.1947\)](#)

[Next issue](#)

Publishing details

Praha: Všichni vydavatelé, 1938-1948

Praha: E. F. Burian, 1938-1941

Praha: Divadlo D 47, 1946-1947

Praha: Divadlo D 48, 1947-1948

Document type

[Newspaper](#)

Keywords

[České časopisy](#)

[Divadelní časopisy](#)

[Měsíčníky](#)

Language

[Czech](#)

Location

[Municipal Library of Prague](#)

Shelf/Call number: F 2335/1938-1939, F

2335/1940-1941, F 2335/1946-1947, F

2335/1947-1948

Keyword Search vs. Semantic Search

	Keyword (Full-Text) Search	Semantic Search
How it works	Matches exact words in content	Uses AI (vector embeddings, NLP) to understand query meaning
Pros	Familiar and predictable Transparent (highlighted terms show why result matched)	Understands context and phrasing Finds conceptually similar content even with different wording
Cons	No understanding of meaning Can return unrelated pages when keywords are scattered	May overlook exact keywords, names, jargon Harder to explain why a result was retrieved

Memory of Newspapers



www.pametnovin.cz



Smart search in historical newspaper archives

Let AI compile answers based on historical newspaper articles from **1880 to 1914** and discover hidden connections to the past

Ask anything...



From:

01.01.1880



To:

31.12.1914



Source

All newspapers



What gifts were given at Christmas?

How did women's suffrage develop?

At what age did women/men usually get married?

Did people complain about the youth?

Which beer was the most popular?

Could women wear pants?

Did people have a problem with alcohol?

How did people spend their free time?

www.pametnovin.cz



Did people complain about the youth?

Response

Yes, people complained about the youth. In an article from ① It is written that the youth in the Ostrava region commit various misdeeds, riots and vices. Similarly, in ② It is mentioned that youth in some districts of Prague and the suburbs are unruly, shouting and shouting, which is considered a problem. Article from ③ He also mentions complaints about young gamblers, smokers, and the obscene language of youth, suggesting that complaints about the behavior of young people were common. In ④ It is stated that youth outside of school are left to their own devices, which leads to misbehavior and disobedience, and that there are constant longings for dissolute youth.



All resources ▾

Resources

① **Ostrava Region**
24.04.1909

Youth is the reserve; when the old fighters for religious, national, cultural and social rights weake...

② **Catholic letters**
28.09.1900

* Youth misconduct. We are told: Especially in some districts of Prague, as well as in the...

③ **National policy**
8.3.1900

*By young people were attracted to them. Gradually, a reading room and a library could also b...

④ **Guild**
06.07.1906

Neglected children. Many social evils could be prevented and prevented if...



⑤ **People's.**
24.12.1910

permanent re the establishr and efforts. w have set then the task of re

Guild

06.07.1906 , Prague

Neglected children. Many social evils could be prevented and prevented if they were nipped in the bud. It is difficult to bend a grown sapling. If we want to have an orderly society, let us educate the youth to be proper people. However, in this respect we see many and many shortcomings in Prague and neighboring cities. Outside of school, the youth are left to themselves, no one cares about their entertainment, few people care about their relationships. What is happening in this matter is insignificant. Our streets are teeming with unruly and disobedient youth, who are therefore such that their parents cannot supervise them, and an institution that would do so and, if necessary, encourage and force the parents to do so, is not such an institution. That is why we hear constant laments and read complaints in the newspapers about unruly and dissolute youth.

Zanedbané děti.

Mnohému společenskému zlu by se předešlo a zabránilo, kdyby zamezovalo se v zárodku. Vyrostlý stromek těžko ovládati. Chceme-li mítí spořádanou společnost, vychovávejme mládež za řádné lidi. V tomto směru však vidíme v Praze a v městech sousedních mnohé a mnohé nedostatky. Mládež mimo školu ponechána jest sama sobě, nikdo nestará se o její zbabavy, mládíko pečuje o její styky.

Co se děje v této věci, jest nezapíraté. Učce naše hemží se mládeží nezvedenou a neposlušnou, která proto jest takovou, že rodičové její na ni nedobhlédnou a zřízení, které by tak činilo a po příslušné rodice k tomu pobádalo i domucovalo, takového zřízení není. Proto slyšíme neustále stesky a čteme v novinách stížnosti na nezvedenou a spustlou mládež.

Bohemia: Catholic political weekly . Prague: Antonín Schmitt, 06.07.1906, 31 (184), p. [1].

Visit the source

věra velmi špatná. Za hrozného sírocco přelodil jen se lilo.

— Z Medkoviče, 6. prosince. (T. k.) Voda stojí 340 centimetrů nad normálem. Sitace jest uspokojivá.

— Z Pešti, 6. prosince. (T. k.) Dunaj v horšější části až k Pešti opadáva. Od Pešti k Zemuni voda stoupá.

** Neštěstí ve slevárně. Z Berlína, 6. prosince. (T. k.) „Voss. Ztg.“ oznamuje z Dobrosoli, že tam v jedné slevárně praskla forma na 130 centů tekutého kovu, při čemž bylo 15 dělníků zraněno, jeden z nich smrtelně.

** Neštěstí na moři. Z Berlína, 6. prosince. (T. k.) „Voss. Ztg.“ oznamuje z Hamburku: Následkem silné vánice a mlhy v Severním moři udřelo se několik nehod. Parník „Alvine“ potopil se na cestě do Rotterdamu a celou posádkou. Loď „Leading Light“ naplnila se vodou, při čemž 4 osoby zahynuly.

** Povodeň v Benátkách. Z Benátek, 6. prosince. (T. k.) Vlny hnáné neobyčejně prudkým větrem zaplavily velikou část města. K polední byl rach v ulicích nemožný. Gondoly zaháňány byly na náměstí sv. Marka. Povodeň poskytovala pohled, jaký už po mnoho roků nebyl k spatření.

— Z Tereva, 7. prosince. (T. k.) Následkem prudké vlny zaplavila okolní 1 m.

Spolkový věstník.

** Z Umělecké Besedy. Výroční valná hromada Umělecké Besedy koná se dne 14. t. m. o 7. hodině večerní ve velkém Svatováclavské založay v Praze. Na pořadu jsou: 1. Zprávy funkcionářů. 2. Doplnovací volby výboru starostovi (p. Fr. Heritesovi, spisovatel v Praze).

** Neštěstí na moři. Z Berlína, 6. prosince. (T. k.) „Voss. Ztg.“ oznamuje z Dobrosoli, že tam v jedné slevárně praskla forma na 130 centů tekutého kovu, při čemž bylo 15 dělníků zraněno, jeden z nich smrtelně.

** Neštěstí na moři. Z Berlína, 6. prosince. (T. k.) „Voss. Ztg.“ oznamuje z Hamburku: Následkem silné vánice a mlhy v Severním moři udřelo se několik nehod. Parník „Alvine“ potopil se na cestě do Rotterdamu a celou posádkou. Loď „Leading Light“ naplnila se vodou, při čemž 4 osoby zahynuly.

** Křesťancko-sociální vzdělávací a pedagogický spolek českých žen a dívek koná dne 13. prosince t. r. o půl 4. hod. členskou schůzi a dne 24. ledna 1904 včeneček — obě v sálech bývalé Svatováclavské založky.

** Z výboru Jednoty občanů pražských. Pražský klub pořádá dne 10. t. m. o 7. hodině večerní v místnostech klubovních k účtení památky 86tých narozenin dra. Fr. L. Biegra předčasný večer, při něm slavnostní řeči pronese pan vrchní říditel dr. Matuší a pan říditel Frant. Šubert. K večeru (tomuto, jenž má býti tichou, ale důstojnou oslavou zemulého náčelníka, ve domácí radě Pražského klubu pp. členy Jednoty občanů pražských. Výbor Jednoty žádá tudíž pány členy, by četnou účastí projevíli souhlas s touto oslavou.

Literatura.

Rádec duchovní. V č. 2. přineseme mimo oznámené práce prof. Žáka snění hřbitovního řádu, jak byl přijat na konferenci skupiny slánské, který mezitím došel schválení vd. konsistoře a sl. c. k. hejtmantví. Pro další číslo připravena osnova stanov Spolku křesťanských matek, kterou i nynější Pius X. doporučil Lev XIII. a kterou i nynější Pius X. doporučil. Též o promluvy, přiměřené jednotám církevním, bude podáno ráno, čímž duchovním správcům bude dána rukověť k vedení podobných spolků na osadách, v nichž jsou podány budou obsahy různých dopisů pastoračních.

Divadlo.

Narodní divadlo. Dnes nastoupí v novém nastudování historické drama „Syn člověka“, od F. v. Jeřábka (4. čtvrt předpl.). Zítra hraje se dvakrát: odpoledne Griarova „Dobrou noc, pane Pantalone!“ a večer pak Charpentierova „Věc z předlohu“.

Literatura.

Rádec duchovní. V č. 2. přineseme mimo oznámené práce prof. Žáka snění hřbitovního řádu, jak byl přijat na konferenci skupiny slánské, který mezitím došel schválení vd. konsistoře a sl. c. k. hejtmantví. Pro další číslo připravena osnova stanov Spolku křesťanských matek, kterou i nynější Pius X. doporučil Lev XIII. a kterou i nynější Pius X. doporučil. Též o promluvy, přiměřené jednotám církevním, bude podáno ráno, čímž duchovním správcům bude dána rukověť k vedení podobných spolků na osadách, v nichž jsou podány budou obsahy různých dopisů pastoračních.

www.pametnovin.cz



10 000 000

Vectors

500 000

Pages

65 000

Issues

1880–1914

Time period

25

Periodicals

5

Libraries

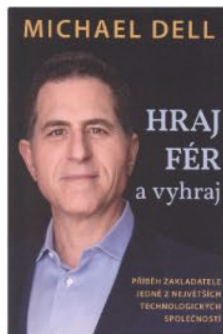
Hlídka času
Ženský list
Ostravský kraj
Hospodářský list
Pražský Merkur
Večer
Jihočeské ohlasy
České slovo
Nový jihočeský dělník
Katolické listy
Plzeňský obzor
Rovnost
Čas
Právo lidu
Brněnské noviny
Čech
Venkov
Našinec
Pokrokové listy
Severočeský dělník
Obuvník
Lidové noviny
Národní politika
Pražské hospodářské noviny

Rating

- exceptionally great: 9+
- great: 8+
- very good: 7+
- good: 6+
- all

Year of release

Of: Do:



Play fair and win.

The story of the founder of one of the largest technology companies

Michael Dell • 2022

The story of the founder and CEO of one of America's largest technology companies. Play Nice But Win is more than just a candid portrait of a leader at a crossroads; it's a survival story that proves that while anyone with technological insight and entrepreneurial drive can build something great, it takes real people... more

[Q Search for similar books](#)

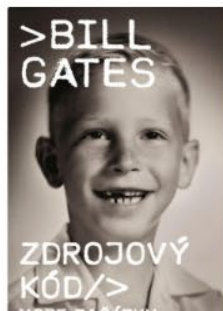
autobiographical reminiscences

autobiographical memories

Leadership

+ 33 more

Exceptionally great **9.0**
2 reviews



Source code/>

My beginnings

Bill Gates • 2025

Source Code is not about the golden age of Microsoft, the founding of the Gates Foundation, or the future of technology. It tells the human and personal story of how Bill Gates became who he is today: his childhood, his early passions, and his hobbies. It is the story of his principled grandmother, his ambitious parents, his first deep friendships, and the sudden... more

Exceptionally great **10.0**
1 review

Memory of Newspapers

- A prototype for semantic search over Czech historical newspapers (1880–1914)
- Users ask questions in natural language. The system finds relevant articles by meaning, not exact words
- An LLM composes answers from top matches, including references and links to original sources
- Retrieves information despite OCR errors or archaic terminology by matching meaning
- Allows filtering by time period and newspaper title

Discover a Book

- A prototype for semantic search over a book catalog
- Covers over one million books
- Uses bibliographic metadata and annotations to capture meaning
- Users ask questions in natural language; the system finds relevant books by meaning, not exact words

Integrating Semantic Search into Digital Libraries

- Challenges:
 - In "Memory of Newspapers": Clearly defined, homogeneous data from a specific period (1880–1914); users motivated to query in natural language.
 - In digital libraries: Diverse content including fiction, old newspapers, contemporary scientific articles, etc.; users accustomed to classic full-text search.
 - "Discover a Book": Focuses solely on books; searches for the book itself, not specific parts of the text within it.
- Main issue: Primarily a UX/UI problem.
- Approach: Separate technology testing from final integration into the digital library UI.
- Prototypes: Built over clear, defined data sets to test technologies independently.

Next Steps

- Current limitation: Both prototypes ("Memory of Newspapers" and "Discover a Book") focus solely on semantic search
- Goal: Build a robust search service by combining multiple approaches for improved accuracy and flexibility
- Key strategies:
 - Hybrid Search: Combine full-text search with semantic search for broader and more precise results.
 - Reranking:
 - Vector-based search efficiently processes millions of records but is not optimal
 - Solution: Use vector and full-text search to identify top N (e.g., top 100) relevant results, then apply slower, more accurate reranking algorithms to refine and filter the results.

Next Steps

- Key strategies:
 - Named entity recognition (NER):
 - Extracts named entities (e.g., proper names, geographic locations, dates) from data and user queries.
 - Automatically applies named entities from user queries as filters to refine search results.
 - Query analysis:
 - Determines whether full-text or semantic search is more suitable for the query.
 - Identifies query components, including language model instructions, that are not suitable for semantic search.
 - Automatically extracts data from queries to use as filters for enhanced precision.

Thank you!



Petr Žabička

petr.zabicka@mzk.cz

www.mzk.cz



Jan Rychtář

jan.rychtar@trinera.cz

www.trinera.cz