Short-Answer Questions

Question 1

Investigate two search engines (e.g., Google and Bing), and figure out what tricks they are using by explaining your testing method. (15%)

a) *Do both search engines use stemming?*

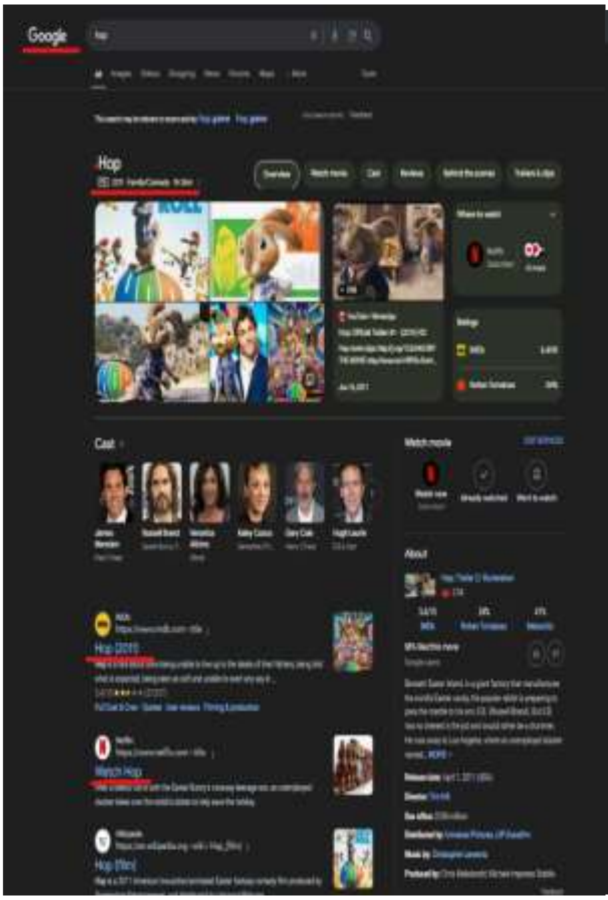
Stemming is a process used by search engines to reduce words to their base or "root" form so that variations of the word are treated the same. For instance, words like “jumpping,” “jumps,” and “jumper” can all be reduced to “jump.” This helps the search engine return more comprehensive results by treating similar words as equivalents.

Testing Method:

1. By performing searches on both **Google** and **Yahoo** using variations of a word I got below results:

    1. Hop, hopping, hopper

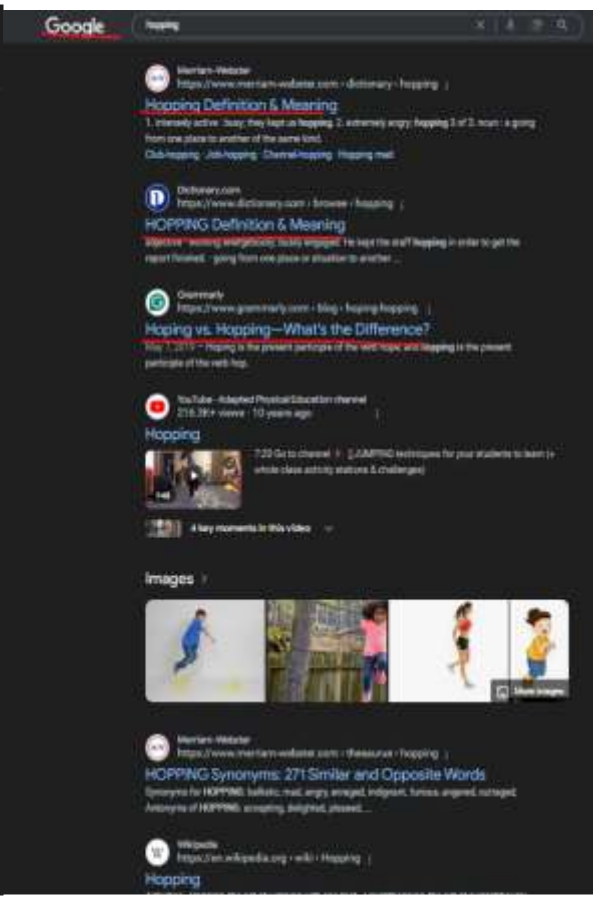
        1. Google-



Google

hop

Hop



Google

hopping

Hopping Definition & Meaning

HOPPING Definition & Meaning

Hoping vs. Hopping—What's the Difference?

Hopping

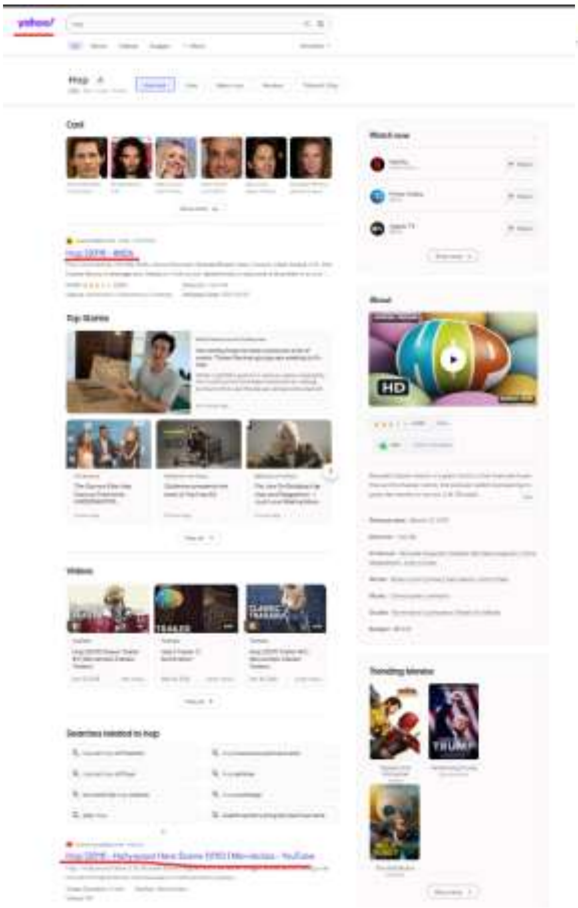
4 key moments in this video

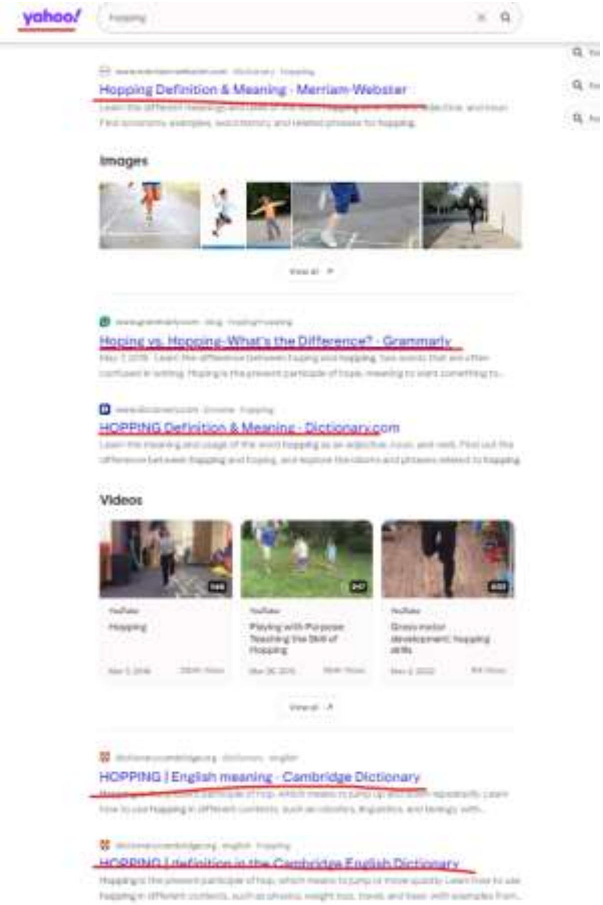
Images

HOPPING Synonyms: 271 Similar and Opposite Words

Hopping

2. Yahoo -





Hopping Definition & Meaning - Merriam-Webster

Images

Hoping vs. Hopping-What's the Difference? - Grammarly

HOPPING Definition & Meaning - Dictionary.com

Videos

YouTube
Hopping

YouTube
Playing with Purpose Teaching the Skill of Hopping

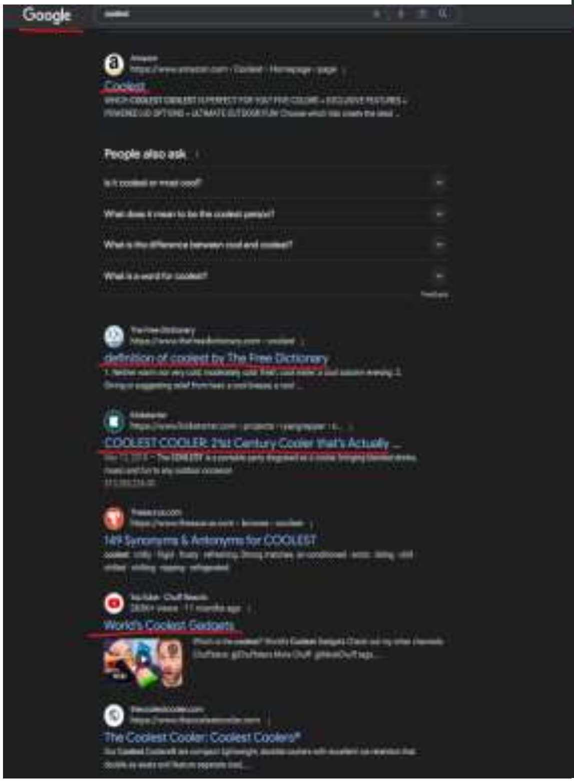
YouTube
Gross motor development: hopping skills

HOPPING | English meaning - Cambridge Dictionary

HOPPING | definition in the Cambridge English Dictionary

2. Cool, cooler, coolest

    1. Google –



Google

Coolest

People also ask

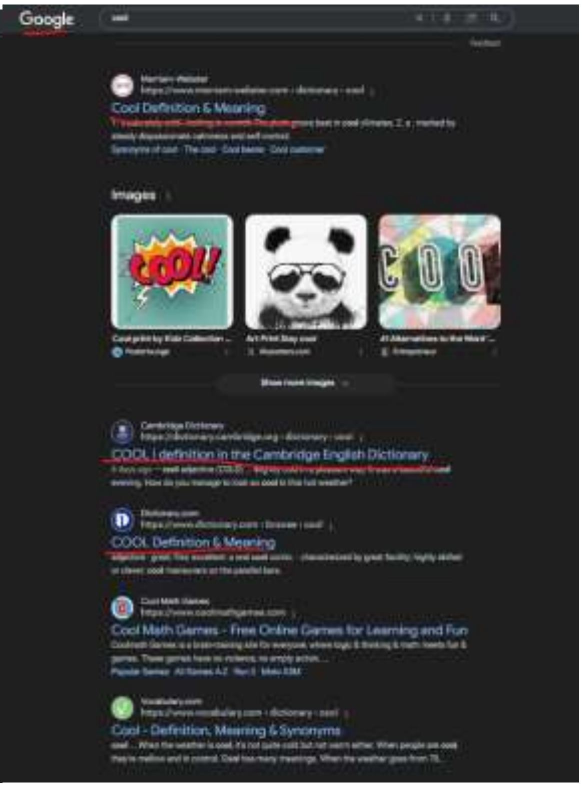
definition of coolest by The Free Dictionary

COOLEST COOLER: 21st Century Cooler that's Actually ...

149 Synonyms & Antonyms for COOLEST

World's Coolest Gadgets

The Coolest Cooler: Coolest Coolers®



Google

Cool Definition & Meaning

Images

Show more images

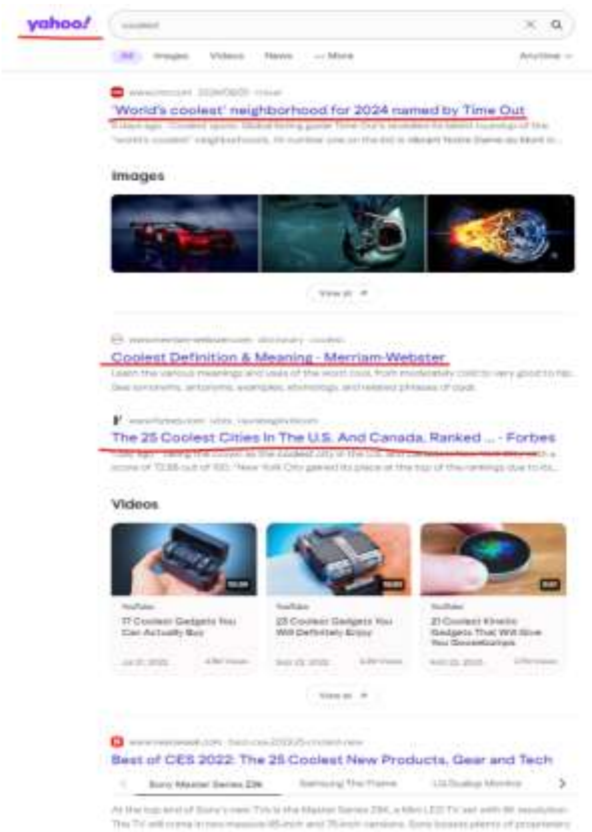
COOL | definition in the Cambridge English Dictionary

COOL Definition & Meaning

Cool Math Games - Free Online Games for Learning and Fun

Cool - Definition, Meaning & Synonyms

    2. Yahoo –



yahoo!

'World's coolest' neighborhood for 2024 named by Time Out

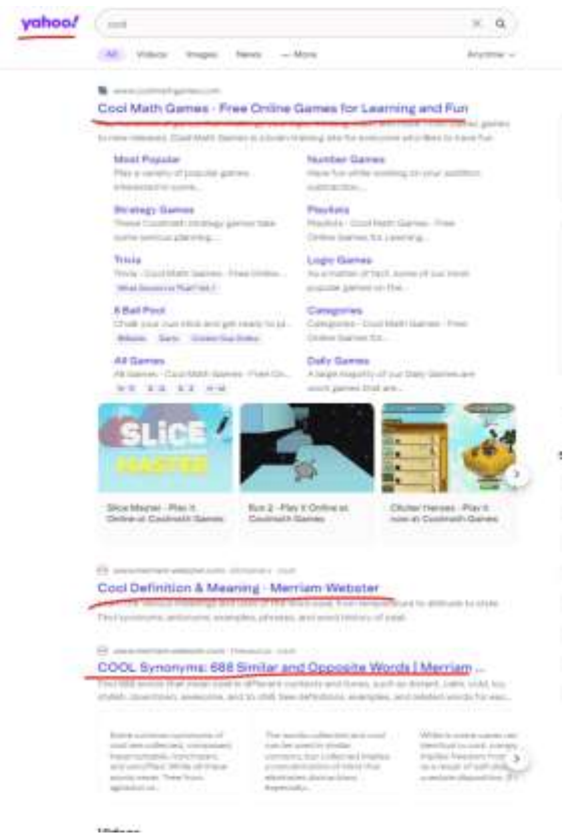
Images

Coolest Definition & Meaning - Merriam-Webster

The 25 Coolest Cities In The U.S. And Canada, Ranked ... - Forbes

Videos

Best of CES 2022: The 25 Coolest New Products, Gear and Tech



yahoo!

Cool Math Games - Free Online Games for Learning and Fun

Most Popular

Number Games

Strategy Games

Playlists

Trivia

Logic Games

8 Ball Pool

Categories

All Games

Daily Games

Cool Definition & Meaning - Merriam Webster

COOL Synonyms: 688 Similar and Opposite Words | Merriam ...

2. The goal was to see if both engines returned similar results regardless of the word variation used, which would indicate they were using stemming or a similar technique.

Findings:

- Both **Google** and **Yahoo** seem to use a more advanced form of word reduction called **lemmatization** rather than basic stemming. Lemmatization goes beyond simply chopping off endings; it looks at the context and meaning of a word.
- For example, searching for "cool" and "coolest" may not yield identical results but will still return relevant content because the search engine understands that "cool" and "coolest" share the same root meaning, rather than just chopping off suffixes mechanically like stemming does.

Conclusion:
While both search engines seem to perform some sort of term normalization, they do not strictly use stemming. Instead, they utilize more sophisticated techniques like lemmatization, which provides results based on the meaning of the word rather than just its root form.

b) Do both search engines filter stop words?

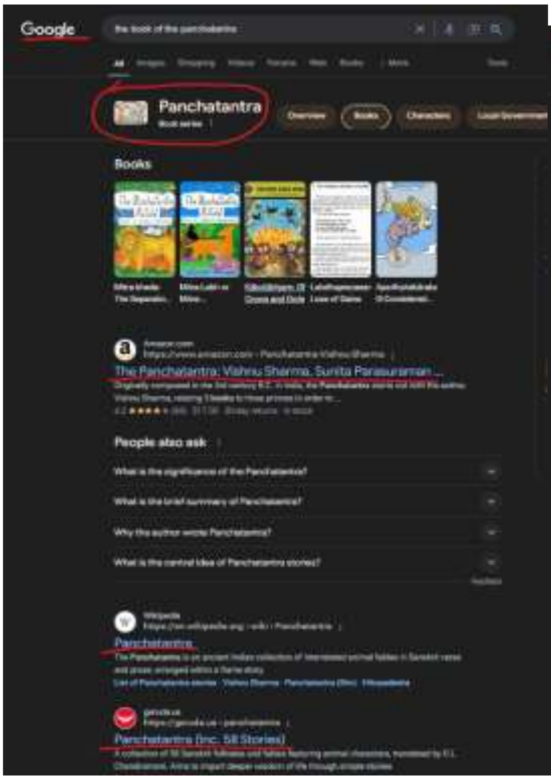
- **Stop words** are common words such as "the," "is," "at," "in," and "of" that typically don't add significant meaning to search queries. Search engines like **Google, Yahoo, Bing, etc.** often ignore these words to speed up searches and focus on the most important terms.

Testing Method:

1. Performing searches with stop words and without stop words on both **Google** and **Yahoo**

    1. "the book of the panchatantra" vs "history world".

        1. Google –



Panchatantra
Book series

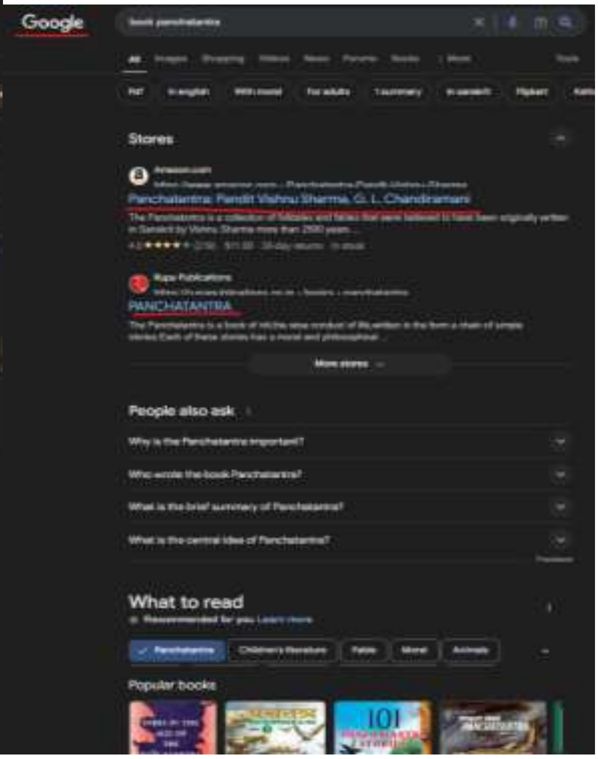
Books

Amazon.com
https://www.amazon.com › Panchatantra-Vishnu-Sharma
The Panchatantra: Vishnu Sharma, Sunita Parasuraman ...

People also ask
What is the significance of the Panchatantra?
What is the brief summary of Panchatantra?
Why the author wrote Panchatantra?
What is the central idea of Panchatantra stories?

Wikipedia
https://en.wikipedia.org › wiki › Panchatantra
Panchatantra
The Panchatantra is an ancient Indian collection of interrelated animal fables in Sanskrit verse and prose, arranged within a frame story.

Panchatantra (inc. 58 Stories)



Stores

Amazon.com
Panchatantra: Pandit Vishnu Sharma, G. L. Chandiramani

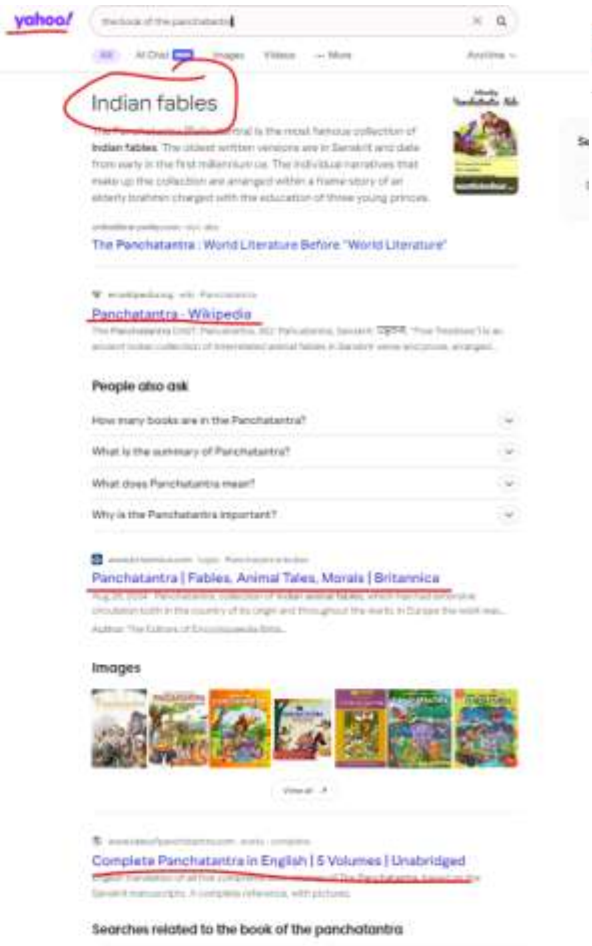
Rupa Publications
PANCHATANTRA

People also ask
Why is the Panchatantra important?
Who wrote the book Panchatantra?
What is the brief summary of Panchatantra?
What is the central idea of Panchatantra?

What to read
Recommended for you Learn more

Popular books

        2. Yahoo -



the book of the panchatantra

Indian fables

The Panchatantra : World Literature Before "World Literature"

Panchatantra - Wikipedia

People also ask

How many books are in the Panchatantra?

What is the summary of Panchatantra?

What does Panchatantra mean?

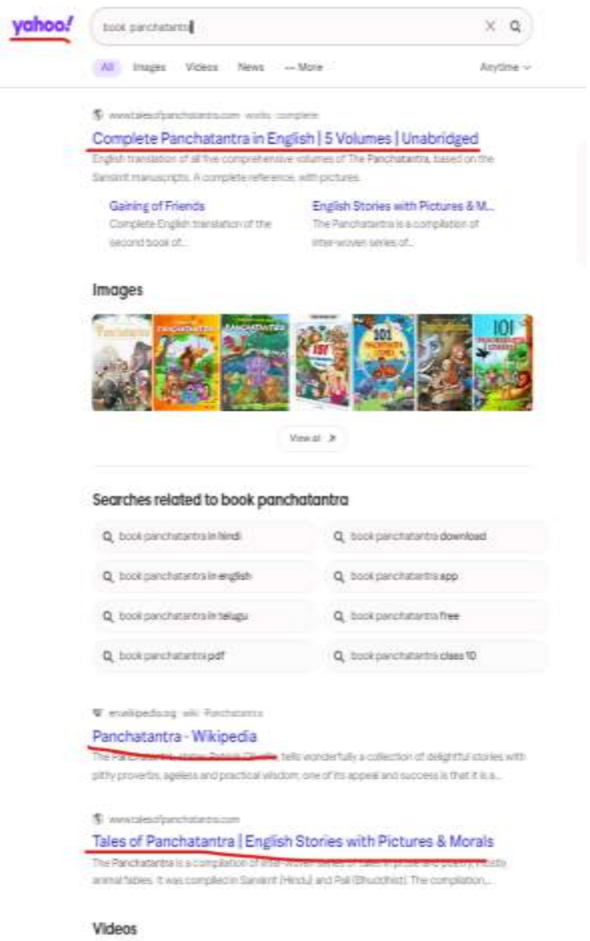
Why is the Panchatantra important?

Panchatantra | Fables, Animal Tales, Morals | Britannica

Images

Complete Panchatantra in English | 5 Volumes | Unabridged

Searches related to the book of the panchatantra



book panchatantra

Complete Panchatantra in English | 5 Volumes | Unabridged

English translation of all five comprehensive volumes of The Panchatantra, based on the Sanskrit manuscripts. A complete reference, with pictures.

Gaining of Friends

Complete English translation of the second book of...

English Stories with Pictures & M...

The Panchatantra is a compilation of inter-woven series of...

Images

Searches related to book panchatantra

book panchatantra in hindi	book panchatantra download
book panchatantra in english	book panchatantra app
book panchatantra in telugu	book panchatantra free
book panchatantra pdf	book panchatantra class 10

Panchatantra - Wikipedia

...tells wonderfully a collection of delightful stories with pithy proverbs, ageless and practical wisdom; one of its appeal and success is that it is a...

Tales of Panchatantra | English Stories with Pictures & Morals

The Panchatantra is a compilation of inter-woven series of tales in prose and poetry, mostly animal fables. It was compiled in Sanskrit (Hindu) and Pali (Bhuddhist). The compilation...

Videos

2. I compared the results to see whether the inclusion of stop words affected the ranking or content of the results.

3. The ranking did differ by a small margin, by prioritising on rank 1 the Amazon(ecommerce website) or the short summary of the book.

Results:

- Both **Google** and **Yahoo** appear to ignore stop words in most cases. For example, searching for " **the book of the panchatantra**" yields the same results as searching for "**book panchatantra**"

- However, there are exceptions. For very specific searches, like exact phrase matches (using quotation marks), stop words do seem to matter. For example, searching for “**The Book**” (with quotes) yields different results than searching for ”**Book**” alone.

Conclusion:
Yes, both search engines filter out stop words in general searches, unless you're doing a very specific search or using quotes for exact matches. In most scenarios, stop words don't affect the results, allowing the engines to focus on the key terms that matter most to the search.

Question 2

For a particular search query, your IR system returns 14 relevant documents and 16 irrelevant documents. There are a total of 80 relevant documents in the collection. (10%)

a) *What is the precision of the system on this search?*

$$Precision = \frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ documents\ retrieved}$$

$$Precision = \frac{14}{14 + 16}$$

$$Precision = \frac{14}{30}$$

$$Precision = 0.4667 \approx 46.67\%$$

The precision of the system for this search is approximately **46.67%**. This indicates that about **46.67%** of the retrieved documents were relevant.

b) *What is the recall of the system on this search?*

$$Recall = \frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ relavent\ documents}$$

$$Recall\ = \frac{14}{80}$$

$$Recall = 0.175\ = 17.5\%$$

The recall for this search is **17.5%**, which means that the system only found about 17.5% of the relevant documents available in the collection, leaving a significant portion undiscovered.