

Practical: Observing how real search engines work

YSE stands for Your (Web) Search Engine. Please choose either Google, Bing or Yahoo!

1. Which engine did you choose?

A: Observing YSE

Type YSE . COM into the location bar of your browser. When the page loads you may find that this has redirected you to a regional variant of YSE, such as YSE . COM . AU . If this happens, navigate around on the YSE page and select the option which says you want to use YSE . COM. Knowing how to do this will help you if find yourself in Brazil but don't understand Portugese! Doing it will also mean we all get the same plain-vanilla experience.

For each of the following experiments, submit a query of the type indicated. If numbers are shown afterward, they provide space for the following observations:

1. How many advertisements are shown?
2. How many “hits” are reported?
3. How many “organic” search results are shown?
4. What results are shown which are not ordinary web pages?
5. How many results would you classify as “website homepages?”
6. Are there any results which don't seem to match the query as you literally typed it? (No need to examine the page.)

A1: Query with the name of a well known brand, e.g. Microsoft, YSE, Sony, Coca Cola, Qantas, etc.

2. ?
3. ?
4. ?
5. ?
6. ?
7. ?

A2: Same query as in A1 but modify it so it is misspelled but the intention is obvious to a human?

1. ?
2. ?
3. ?
4. ?
5. ?
6. ?

A3: Query with the name of a pop culture icon: e.g. Lady Gaga

1. ?
2. ?
3. ?
4. ?
5. ?
6. ?

A4: Try a very specific query: 'Jane Davidson Karen Mattock Jon Drummond Karen Croot Roger Dean Paul McCormack'

1. ?
2. ?
3. ?
4. ?
5. ?
6. ?

A5: Now try introducing spelling errors into words in that query until a “Did you mean?” appears or YSE reports that no results are found.

1. How many words did you need to corrupt?
2. What happened?
3. Did it matter which words you corrupted?

A6: Now repeat the A5 experiment by going back to the original A4 query and inserting a + before each word. E.g '+Jane +Davidson'

1. How many matches do you now get for the uncorrupted queries
2. How many words now need to be corrupted to produce the same behaviour as in A5?

A7: See if you can determine whether YSE treats word variants e.g singular/plural as equivalent. E.g. If you enter 'crickets' will it show you results for 'cricket'? What about other words? If you enter 'cricketer' or 'cricketing' do you get results for insects ... in subsequent pages in the ranking?

A8: Does YSE automatically look for synonyms? Try 'petrol price tennessee'. Does it give you 'gas' prices?

A9: Is YSE behaviour in your A7 and A8 experiments modified if you insert a '+' before the word(s)?

A10: Does YSE query processing depend upon word order? Try queries like 'London School of Economics' and then reverse the word order.

A11: Sometimes multiword queries form a phrase (as in the previous example) but other times a rare combination of words is chosen with the intention to intersect on the wanted page or pages, e.g. 'Sunderland Qantas Indian Ocean'. How effective is YSE at identifying the former intent, i.e. how many of the result pages for queries like A10 actually match the phrase? If you surround the query with double quotation marks:

1. How does the first page of results change?
2. How does the reported number of hits change?

A12: What results do you get if you enter a “stopword” such as 'the'?

1. What characterises the top-ranked results?
2. What is the reported number of hits?
3. Could this number be used to estimate the number of pages in the YSE index?
4. Can you think of a better method of estimating the size of the YSE index based on hit counts?

A13: What do you expect will be the top ranked results for a single-letter query such as 's'? Try it.

1. What did you notice?

A14: Try the query 'U.S.A', then try replacing the dots with other punctuation characters, e.g. 'U-S-A' and 'U/S/A'.

1. Do the results change?

A15: Incidentally, what happens if you change to lower case letters?

1. Do the results change?

A16: How does YSE handle accented letters? Try the query 'Sodertalje' which, when written with the appropriate accents, is the name of a town South of Stockholm.

1. Does the unaccented query appear to match accented versions of the word?

A17: Is YSE's spelling suggestion system like the spelling correction system in a word processor? Try the queries 'Manchester Untied' and 'Collingwood Furball Club'. Then try the query 'the evilla empire'.

1. Does YSE sometimes make 'spelling' suggestions when all the words are correctly spelled?
2. In the last example, conventional spelling suggestion would most likely suggest replacing 'evilla' with 'villa' which is only one edit away. Is that what happens?

A18: Pick a query in an area which is likely to have rapidly changing websites – perhaps related to the World Cup or an unfolding news story.

1. Submit the same query a few times – Do the results change at all?
2. Coordinate with your neighbours to use the same query – can you spot any discrepancy in results?

A19: Try entering a query in mathematical notation, e.g. $P(A|B,C)$

1. What do you observe?

A20: How quickly can YSE update its index? Try the query 'current time Bowral' and check what time is shown in the results summary for one of the website results (not the YSE answers at the top.) Don't open the page itself, as it ought to show the right time.

1. Does the day and time shown in the summary tell you when YSE last visited this page?

A21: Try entering the query 'bank'. Then connect to a regional YSE (e.g. YSE.com.au or YSE.co.uk) and submit it again.

1. Do the results change?
2. What do you conclude?

B. More complex search tasks on YSE

B1. You're planning a bicycling holiday in Kyrgystan. Can you use YSE to find any webpages offering specific advice on health precautions you should take?

B2. You have \$50 million in cash (what are you doing here?) and would like to invest it for high return without unacceptably high risk. Can YSE provide you with all the information you need to guide your investment?

B3. You need to document occasions where a U.S Secretary of State or Secretary of Defense has met with a foreign leader to discuss the provision of any type of weapon of mass destruction. Can YSE provide you with a list?

B4. I'm having trouble booting my Sony Z laptop under Kubuntu 10.04. Is it due to the RAID controller or to the hybrid graphics system? Why can't I boot with the VGA cable connected? Help!

C. Non-YSE search

C1. Navigate to nps.org.au Enter the query 'diabetes'

1. What do you observe about the results?
2. Why do you think results are presented this way?

C2. Navigate to the online catalogue of a library you know. Can you locate the French version of the book which in English is called *the Outsider*, by Albert Camus?

1. How did your query differ from the ones you used on YSE?
2. How did the results differ?

C3. Navigate to australia.gov.au and search for 'recycling'.

1. What do you observe about the results?
2. Why do you think results are presented this way?
3. How do you think the query suggestions on the right are generated?

C4. Enter a postcode into the postcode box and resubmit the same query.

1. What do you observe about the results?
2. How do you think this mechanism works?

C5. Navigate to clivepeeters.com.au and search for 'television'. Use the facet links on the left to repeatedly narrow down to a particular television you might buy.

1. What are the advantages of this type of searching/browsing?