

WolframAlpha – How it Works? [Part 2]

Launched on 18th May, the long-term goal of Wolfram|Alpha ('WA' from now on) "is to make all systematic knowledge immediately computable and accessible to everyone".

In Stephen Wolfram's own words, "when computers were young, people assumed that they'd be able to ask a computer any factual question, and have it compute the answer." I'm sure Wolfram would have like to have added "and I'm happy to say that ..."

If WA's goal is to deliver such a system; well, it would seem that it's the tool many of us have been dreaming of!

In my previous article I tried – as have many – to say something about how the system *appears* to function, and in this – the round up – I'll take that a little further, and tell you why I'm a little disappointed at the moment – yet hopeful too.

From part 1, I suggested that WA works by a) feeding your query through a linguistic-parser, b) taking that output, and applying further rules and manipulation, and, c) assessing deep-webⁱ data to provide – hopefully – an answer.

One of the hardest things about WA is, um, using it. That is to say, getting your query 'correct'; or at least in a form that WA can understand and use.

Understand and Use

'understand' means (these are my definitions), that the linguistic-parser (all the parts: syntactic, semantic, etc) has made some initial 'sense' of your words; whereas 'use' implies that this resultant 'sense' is able to be further understood and manipulated in order to be mapped to WA's deep-web data-sources.

Important point: in order to get a successful result, you must realise that although WA might *like* your query (all but one of the criteria above were satisfied), it might still be short on relevant data. However, that *just* requires adding more structured data sources! Worth bearing in mind though.

Getting a successful result also means you've not misunderstood WA's function altogether [repeat after me 'Wolfram Alpha isn't Google', 'Wolfram Alpha isn't Google' ...]

WA has some useful (yet minimal) guidance here which highlights the differences:

Important things to know about Wolfram|Alpha:

- **Wolfram|Alpha answers specific questions rather than explaining general topics**
Enter "2 cups of sugar", not "nutrition information"
- **You can only get answers about objective facts**
Try "highest mountain", not "most beautiful painting"
- **Only what is known is known to Wolfram|Alpha**
Ask "how many men in Mauritania", not "how many monsters in Loch Ness"
- **Only public information is available**
Request "GDP of France", not "home phone of Michael Jordan"

If Wolfram|Alpha is still not sure what to do, try the following:

- **Don't use long complete sentences; just enter the minimum number of words needed to communicate**
- **Try different words or notations**
- **Use whole words instead of abbreviations**
- **Check your spelling**

» [Ask the Wolfram|Alpha Community for Help](#)
» [Examples by Topic](#) » [Quick Video Overview](#)

If you enter the first-parts of these examples into WA you'll mostly get 'proper answers', whereas you won't with the second bits ("highest mountain" doesn't actually produce any output. Oops!) With Google, either part results in 'hits'; but that's what you'd expect – however, don't expect to get 'proper answers', and in fact, be surprised if you do!

So one hurdle in using WA effectively is to get to grips with how to create, and ask a good question.

I can think of one easy way to improve learning this skill perhaps: give users the chance to 'rate' their WA answers: "yes, that's exactly what I wanted!"; "WTF!"; "oh no - not that 'WA isn't sure what to do with your input' message again!" Or perhaps just a simple 1 to 10 score!

Then, as a learning aid, WA could display lists of previous queries that scored highly - what better way to learn than to see what's previously worked (and, of course, ask yourself 'why anybody would want to know *THAT!*')?

Anyway, let's have a look at an actual attempt to use WA; one where WA produces data that's at least semi-relevant.

The problem I wanted answering was to do with tossing a coin multiple times - yet with a twist in the tail.

So to start with I entered 'coin toss'.


WA came back with some basic probability stuff on tossing a coin - like the probability of seeing 12 heads and 8 tails given 20 tosses.

However, the question I wanted answering was quite a bit more complex (that's why I *want* this tool, i.e., I'll do the easy stuff!), and is to do with tossing a coin, **and** seeing a sequence in the results. Here's the question:

Q. Given a series of consecutive coin tosses, is it more/less likely that I'd see the sequence HTT before I saw the sequence HTH - or is it the same?

Just to cut to the chase here; the answer is that the average number of tosses needed to see HTH is 10, whereas it's just 8 for HTT (if you don't believe me, try it! Or just email me for an explanation).

Ok, so on entering that straight into WA I got what you might expect:

 *Wolfram|Alpha isn't sure what to do with your input.*

And that's pretty reasonable, isn't it?

For one thing, I actually re-typed that question about ten times before I was pretty sure that I'd made the problem pretty clear, to *you*, the reader (you'll be the judge of course!); and all linguists know that I could re-word that question thousands of times, yet still have it make just about as much (or better) sense. So, the "Wolfram|Alpha isn't sure what to do with your input" message might be a linguistic failure - to 'understand'?

However, maybe it did well with that - after all, I gave it a fair few hints: 'coins', 'tosses', 'likely', 'sequence', 'T', 'H' - and we've already established that WA does know stuff about tossing coins!

I spent a good half an hour trying various ways of asking this question, but try as I might, I couldn't get any further on it (so you're going to have to try it for yourselves), and this might be simply because WA hasn't enough relevant data to draw upon yet ... OR, as I hope I've made clear already, that it can't extract enough 'clues' about what I'm asking from its linguistic analysis of the input ... OR, that it can't take adequate clues, and then manipulate them in such a way as to make querying its data-sources work in viable way. Who knows?

Perhaps another improvement would be to have WA tell you more about its 'I'm not sure' message? What *exactly* was the problem?

In the end I found myself muttering, 'come on, it's only a particular case of conditional probability', and permutations, and so I found myself trying 'conditional probability' as a potential new starting-point.



The screenshot shows the WolframAlpha search interface. At the top, the WolframAlpha logo is displayed with the tagline 'computational.. knowledge engine'. Below the logo is a search bar containing the text 'conditional probability'. The search results are displayed in a white box with a red header 'Logical Fallacies'. Below the header, there is a message: 'Development of this topic is under investigation...' accompanied by a gear icon. Below this message is a form with the text 'Leave your email address to show your interest.' and a 'send' button. Below the form, there is a section titled 'More to explore:' with a grid of links: Physics, Life Sciences, Transportation, Education, Technological World, Mathematics, Food & Nutrition, Materials, and More... At the bottom of the results box, there is a link: 'Are you an expert on this topic? Find out how you can help. »'

Well it seems as though I'll have to wait, and, in my gut, I think I'll probably have to wait a fairly long time before I can enter something along the lines of my question - and see something pertinent come back.

Ok, so it **was** a 'tough ask' - I acknowledge that - and I'll also admit that I don't profess to know all the ins/outs of how to interact with WA. Should one *have* to know this though? If WA is hoping to be a (Natural Language) Computational Knowledge Engine - which I assume **is** the ultimate goal - it has quite some way to go - unless you want to know how high Everest is in terms of Golden Gate bridges of course!

ⁱ Deep-web data usually means data that is accessible, if only you knew how to get at it, and that's not quite right here, i.e., the data WA uses might include stuff that's only available via specialist data-feeds etc.