

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



Wolfram|Alpha

ESEJ DO PŘEDMĚTU DIGITÁLNÍ KNIHOVNY

Jiří Mauritz, 409972
3. ročník Bakalářského studia

Brno, 2. prosince 2014

Informace o projektu

Projekt *Wolfram/Alpha* je výpočetní vědomostní systém¹, který odpovídá na širokou škálu dotazů na základě svých uložených a zpracovaných dat. Projekt je vyvíjen společností *Wolfram Research* v čele se zakladatelem Stephenem Wolframem. Jako webová služba je dostupný na stránkách <http://www.wolframalpha.com/> od 18. května 2009.

Úvod

Wolfram/Alpha se liší od ostatních vyhledávačů tím, že místo aby uživatele odkazoval na jiné stránky, snaží se na otázku odpovědět sám z informací, které má k dispozici. Jeho další předností je kvalita zpracování přirozeného jazyka. Přestože vyhledávač zatím odpovídá pouze na anglicky položené dotazy, můžete se ho zeptat rozvitou otázkou jakoby jste se ptali lidského protějšku. Odpověď je stejně bohatá jako může být otázka – kromě výsledku systém vypíše a vykreslí mnoho dalších informací souvisejících s výsledkem. To je jeden ze znaků, že se vývojáři zaměřují na vzdělávací funkci. Pokrytí možných otázek má samozřejmě svou hranici. *Wolfram/Alpha* odpovídá jen na faktograficky založené otázky, na které je jasná objektivní odpověď, například: „*What was the weather in London when Michael Jackson was born*“. Na otázky typu „*How does car work*“ nebo „*Who is more pretty*“ vrátí fakta o pojmech, které najde v otázce, ale na odpověď přímo neodpovídá.

Vývoj

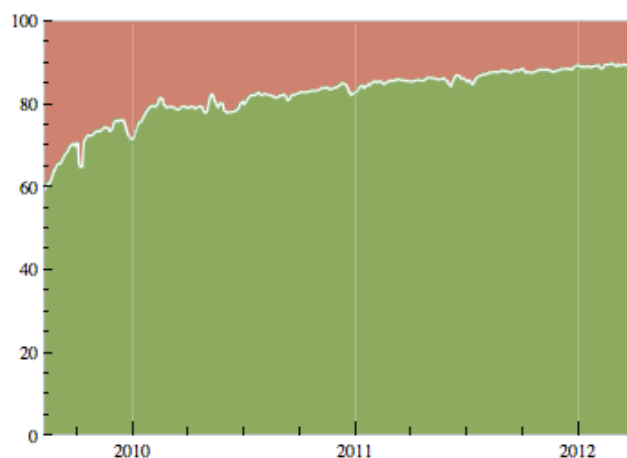
Jádro projektu *Wolfram/Alpha* je postavené na výpočetním systému *Mathematica*, který vznikl mnohem dříve, než byl *Wolfram/Alpha* spuštěn. Proto se dá za počátek jeho vývoje považovat rok 1988, kdy Stephen Wolfram vydal první verzi softwaru *Mathematica*. Obsahovala jen asi 600 funkcí oproti dnešním 2500, ale pro svět matematiky a fyziky představovala *Mathematica* velký pokrok [1]. Zvláště propracovaná funkce integrování vzbudila největší obdiv. Od té doby se *Mathematica* stále zdokonaluje a v současné době je propojená s *Wolfram/Alpha*, takže v ní lze počítat s informacemi z databáze znalostí. Pro takovou komunikaci, která přenáší data v podobě symbolických matematických formulí se používá protokol MathLink.

1. z ang. computational knowledge engine

Zde je přehled funkcí, které byly do *Wolfram/Alpha* přidány od jeho spuštění.

2009	<ul style="list-style-type: none">• statistické modely• medicína (DNA)• mezinárodní měřítka kvality života• telefonní čísla• data o audio průmyslu
2010	<ul style="list-style-type: none">• informace o složení jídel a léků• analýza fyzických aktivit
2011	<ul style="list-style-type: none">• generování QR kódů• sledování fází Měsíce• sledování tras letadel
2012	<ul style="list-style-type: none">• analyzování vašeho Facebook profilu• sledování Mezinárodní vesmírné stanice• rozšířené step-by-step řešení matematických příkladů• hledání nejbližších míst (banky, restaurace atd.)• informace o justičním systému
2013	<ul style="list-style-type: none">• ještě větší zaměření na výuku s Problem generator• práce s obrázky - filtry: jak vidí pes nebo barvoslepí lidé
2014	<ul style="list-style-type: none">• práce s jednotkou Bitcoin• nové rozhraní pro chemické výpočty• mnoho dalšího http://www.wolfram.com/mathematica/new-in-10/

V souvislosti se zdokonalováním databáze znalostí se vylepšovala i bilance zodpovězených odpovědí. Pokud na danou otázku *Wolfram/Alpha* nezná odpověď, nezkouší odhadovat, ale informuje vás o tom, že nedokázal interpretovat váš dotaz. Následující graf ukazuje poměr počtu zodpovězených a neúspěšně interpretovaných otázek v průběhu tří let po spuštění. [2]

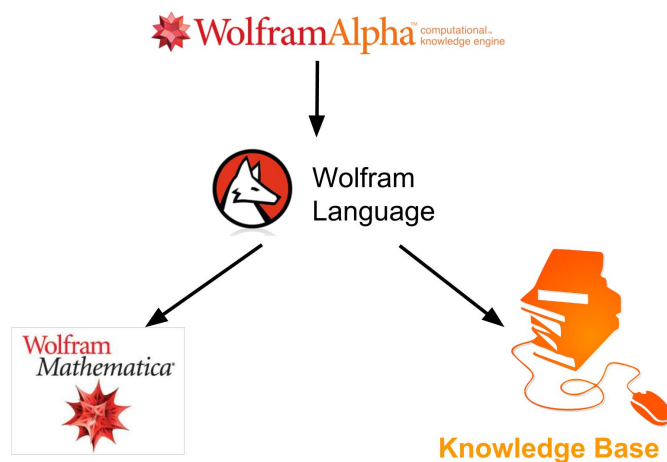


Obrázek 1: úspěšnost dotazů ve *Wolfram/Alpha*

Jak Wolfram|Alpha funguje

Kouzlo úspěšnosti *Wolfram|Alpha* je díky dvěma základním principům práce s daty. Tím prvním je symbolická reprezentace všeho s čím pracuje. *Wolfram|Alpha* umí pracovat s proměnnými, aniž by se staral o to, zda je to číslo, složitá formule nebo jakýkoli jiný objekt. Všechny operace jsou napsány tak, aby mohli pracovat s neznámými a vracet obecné výsledky. Na dotaz $\lim_{x \rightarrow 5} \frac{x}{y^2}$ vám bez namítání vrátí $\frac{5}{y^2}$ a vykreslí funkci v závislosti na y . Po zadání stejné formule doplněnou o $y \rightarrow \infty$ je výsledek 0. [3]

Druhým principem je maximalizace automatizace. Předpokládá, že chcete co nejrychleji výsledek a nezáleží vám na tom, jak výsledek vypočítá. *Wolfram|Alpha* například umí detekovat hrany v obrázku, ale algoritmus detekce je pro nás černá skříňka. Kromě výběru algoritmu pro nás volí vhodnou formu vizualizace, doplňující informace a v neposlední řadě k samotnému porozumění přirozeného jazyka potřebuje nějaký automatický nástroj k uhodnutí sémantiky [4]. Zajímavé je, že tyto principy aplikoval Stephen Wolfram na projekt *Mathematica* v 80 letech a stále jsou přítomny v jeho dalších projektech včetně *Wolfram|Alpha*.



Obrázek 2: zpracování dotazu ve *Wolfram|Alpha*

Wolfram|Alpha je program o 15 milionech řádcích programovacího jazyka *Wolfram Language* běžící na 10 000 procesorech. Tento jazyk slouží jako rozhraní pro zadávání dotazů v softwaru *Mathematica* a je jeho součástí. V červnu 2014 byl však osamostatněn a byl spuštěn *Wolfram Language Cloud* [5], na kterém si můžete vytvořit aplikaci v jazyce *Wolfram Language* s využitím databáze znalostí. Na obrázku 2 je znázorněno zpracování dotazu v systému *Wolfram|Alpha*. *Wolfram Language* zpracuje dotaz pomocí výpočetní síly softwaru *Wolfram Mathematica* a databáze znalostí.

Databáze znalostí

Vývojáři uvádí [6], že je v jejich databázi znalostí zpracovaných několik biliónů elementů. Data byly získány z různých veřejných i soukromých internetových zdrojů. U každé odpovědi je volba *Sources*, která vám odhalí, odkud byla informace získána. Bohužel je ale vždy primárním zdrojem *Wolfram/Alpha Knowledgebase, 201X*, což vám o původu informace neřekne vůbec nic. Většina faktů je získána z těchto zdrojů: World Factbook od CIA, Geologický průzkum Spojených států amerických, knihovnická publikace zvaná All About Birds z Cornell University, biografický slovník Chambers, Dow Jones, the Catalogue of Life [7]. Velmi důležitým zdrojem je i Wikipedia, která je jednak zdrojem „lidových informací“ o každodenní činnosti ve všech oborech a jednak zdroj vyjadřování, díky kterému se *Wolfram/Alpha* trénuje v rozpoznávání přirozeného jazyka. Data prošla skrz 50 000 filtrů, modelů a algoritmů a následně jsou udržovaná a aktualizovaná. Nejdynamičtější fakta (počasí, burzy, ...) jsou aktualizovaná v podstatě neustále. Ty ostatní se obnovují denně, týdně nebo měsíčně podle toho, kdy jsou informace k dispozici.

Databázi znalostí má *Wolfram/Alpha* skutečně obří, ale nelze si nevšimnout, že se zaměřuje na anglicky mluvící země. Skutečně některé zajímavé statistiky jsou velmi omezené pro Českou republiku. Například dotazem „*house prices North Dakota*“ zjistíte statistiku cen nemovitostí v Severní Dakotě, ale dotaz „*house prices Czech republic*“ zůstává nezodpovězen. *Wolfram/Alpha* nemá ani informace o poštovních směrovacích číslech, jen o ZIP kódech. Na druhou stranu má informace o tom, kolik žáku základních škol opakovalo ročník od roku 1990 v České republice. Zdá se, že informace jsou rozmanité a z mnoha zdrojů, které se zaměřují třeba jen na jedno téma.

Závěr

Wolfram/Alpha je bezpochyby revoluční projekt. Když si v minulém století lidé začínaly uvědomovat sílu výpočetní techniky, měli různé sci-fi představy o budoucnosti. Jedna z nich byla o stroji, který by dokázal odpovědět na jakoukoli otázku – podobně jako počítač ze slavného sci-fi seriálu Star Trek. Přestože počítač ze Star Treku zvládl i materializovat libovolnou věc, je mezi ním a *Wolfram/Alpha* jasná souvislost. Osobně si myslím, že jeho největší přínos je v oboru matematiky. Pro matematické výpočty ho pravidelně používám, kdežto na otázky z jiných oborů jsem se ptal jen tehdy, když jsem zkoumal rozsah jeho znalostí. Drtivá většina těchto informací je dohledatelná přes Google nebo na Wikipedii. Aktuální počasí, burzy, historie, statistické informace a mnoho dalších informací je dostupných na internetu. Z tohoto hlediska se nezdá *Wolfram/Alpha* tak převratný. Na druhou stranu vývojáři *Wolfram Research* přidávají stále nové prvky a mají nové nápady (zmiňovaný *Wolfram Language Cloud*). Z hlediska vyhledávání informací *Wolfram/Alpha* pravdě-

podobně nepřekoná Google, ale díky vynikající umělé inteligenci, zpracování přirozeného jazyka, vizualizaci výsledků a hlavně zmiňované zpracování matematických úloh je *Wolfram/Alpha* geniální nástroj.

Bibliografie

- 1 ABBASI, Nasser M. *A little bit of Mathematica history* [online]. [cit. 2014-12-03]. Dostupný z WWW: http://12000.org/my_notes/compare_mathematica/index.htm.
- 2 WOLFRAM, Stephen. *Overcoming Artificial Stupidity: WolframAlpha Blog* [online]. 2014 [cit. 2014-12-03]. Dostupný z WWW: <http://blog.wolframalpha.com/2012/04/17/overcoming-artificial-stupidity/>.
- 3 *Values for Symbols*. [online]. [cit. 2014-12-03]. Dostupný z WWW: <http://reference.wolfram.com/language/tutorial/ValuesForSymbols.html>.
- 4 WOLFRAM, Stephen. *Wolfram|Alpha Is Launching: Made Possible by Mathematica: WolframAlpha Blog* [online]. 2009 [cit. 2014-12-03]. Dostupný z WWW: <http://blog.wolframalpha.com/2009/05/15/wolframalpha-is-launching-made-possible-by-mathematica/>.
- 5 RESEARCH, Wolfram. *Wolfram Programming Cloud: Introducing a programming revolution* [online]. 2014 [cit. 2014-12-06]. Dostupný z WWW: <http://www.wolfram.com/programming-cloud/>.
- 6 RESEARCH, Wolfram. *Frequently Asked Questions: Data in Wolfram|Alpha* [online]. [cit. 2014-12-06]. Dostupný z WWW: <http://www.wolframalpha.com/faqs5.html>.
- 7 JOHNSON, Bobbie. *Where does Wolfram Alpha get its information?* [online]. [cit. 2014-12-06]. Dostupný z WWW: <http://www.theguardian.com/technology/2009/may/21/1>.