

DUM č. 13 v sadě

33. Inf-9 Počítačová lingvistika a typografie

Autor: Roman Hrdlička

Datum: 26.06.2014

Ročník: 3AV

Anotace DUMu: Rozbor nástroje Kontrola pravopisu a gramatiky MS Office.

Materiály jsou určeny pro bezplatné používání pro potřeby výuky a vzdělávání na všech typech škol a školských zařízení. Jakékoliv další využití podléhá autorskému zákonu.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

13. Kontrolor české gramatiky

Než přejdeme k nejdiskutovanějším lingvistickým problémům, je na místě představit nástroj, který již několik verzí umožňuje i českým uživatelům kancelářského balíku Office mít jakousi základní možnost kontroly, zda jimi napsaný text neobsahuje nějaké zásadní nedostatky. Ano, mluvím o nástroji Kontrola pravopisu a gramatiky.

Tento nástroj existoval v podobě Kontroly českého pravopisu již na přelomu tisíciletí, kdy byl schopen na základě slovníků provádět kontrolu překlepů a pravopisných chyb na úrovni jednoho slova. Takové chyby, které znamenají, že slovník dané slovo neobsahuje, takže je pravděpodobné, že obsahuje chybu, jsou nástrojem podtrženy červeně.

Nicméně 30. června 2005 byl poprvé uveden pro Office 2003 také nástroj Kontrola české gramatiky, jenž byl vyvíjen více než 5 let ve spolupráci s Ústavem pro jazyk český Akademie věd České republiky a odborníky z Univerzity Karlovy. Programovou integraci do balíku Office a systému Windows zajistila maďarská společnost MorphoLogic. Tento nástroj má oproti předchozímu vyšší rozsah účinnosti, jelikož umožňuje kontrolu dodržování gramatických pravidel v kontextu celých vět díky propracovanému systému lingvistické teorie pro gramatickou korekci. Tyto gramatické chyby jsou označeny zelenou vlnovkou pod těmi částmi věty, v nichž se podle nástroje vyskytuje chyba.

Jedním z hlavních kritérií při vývoji nástroje přitom byla minimalizace výskytu tzv. falešných poplachů, tedy chybových hlášení poukazujících na jevy, jež jsou ve skutečnosti správně, jinak by jej uživatelé nepoužívali a vypínali. Základem je tedy lingvistické pravidlo „vždy nesprávných vět“. Druhým kritériem je samozřejmě počet reálně odhalených chyb. Nástroj tedy poukáže na ty věty či slovní spojení, které jsou v každém případě a prokazatelně špatně. Textu samozřejmě nemůže „rozumět“, věty hodnotí podle gramatických pravidel na základě rodu, pádu, čísla a dalších mluvnických charakteristik, ve kterých se musejí větné členy shodovat.

Činnost aplikace je možné rozdělit do čtyř kroků:

1. Segmentace textu na slova, čísla, interpunkci a věty.
2. Analýza správnosti věty.
3. Vyhledání a ohlášení chyby.
4. Návrh na opravu.

Pokud se analyzátoru něco nelíbí, podtrhne to zelenou vlnovkou a po kliknutí pravým tlačítkem se uživatel doví, jakou chybu analyzátor našel, jak ji opravit nebo dostane alespoň doporučení, aby věnoval pozornost určitému slovu nebo úseku.

Kontrola české gramatiky si tedy všímá například těchto jevů:

- shody přísudku s podmětem (zaregistruje chybně např. tvar: Děvčata se smály.)
- chybějící nebo přebývající interpunkce (Byl to člověk který se nikdy nevzdal.)
- chyby v zájmenech (Žena, který zvítězila. nebo Udělal to pro jej.)
- chyby ve slovesných tvarech (Chtěl bych, aby jste se tu cítili dobře.)
- chyby v předložkách (Půjdeme na nimi.)

Je ovšem třeba zopakovat, že vzhledem k již zmíněné metodě odhalování chyb, tedy hledání vždy nesprávných konstrukcí a samozřejmě vzhledem k charakteru češtiny jako jazyka, je Kontrola české gramatiky k určitým prohrěškům relativně velkorysá, což může nastat ze dvou různých důvodů.

Za prvé to může být tím, že věta, která člověku připadá chybná, přece jen nějaké formálně správné (byť v reálném světě nesmyslné) čtení má a je tedy vlastně gramaticky správně. Například věta Na dvoře si hráli hezké dívky. je na první pohled špatně, v přísudku by mělo stát „hrály“. Jenže slovo „dívky“ je z formálního hlediska (rod, vzor) úplně stejný tvar podstatného jména jako slovo „hry“. A věta Na dvoře si hráli hezké hry. špatně není, jedná se o nevyjádřený podmět, hry jsou předmětem. A věta Přišel jsme, viděl jsme. je sice z hlediska běžné mluvnice příkladem špatného použití čísla u slovesa, jenže také je v češtině zvláštní variantou tzv. mykání, mluvy používané panovníky, což je sice dnes již záležitost archaická, ale v podstatě zcela správná.

Nu a druhá možnost je situace, kdy z nějakých důvodů bude chyb ve větě tak velké množství, což systém buď zmate, nebo i když vnitřně správně rozpozná, že je věta špatně, mu to znemožní poskytnout uživateli rozumný návrh opravy, takže raději žádnou chybu nehlásí.

Závěrem je třeba ještě podotknout, že v současné verzi analyzátor odhaluje také běžné typografické chyby, jako jsou vícenásobné mezery či chybně zapsané závorky. Metodika hledání chyb byla nejspíš zvolena správně – přestože mnozí spílají tvůrcům, že nestvořili dokonalý nástroj, nelze očekávat s jakýmkoli nárůstem schopností umělé inteligence, že by se lingvistické schopnosti počítače vyhouply na lidskou úroveň. Proto ani sebelepší verze Kontroly gramatiky nemůže nikdy nahradit jazykové vzdělání, k němuž se pokusím přispět dalšími díly těchto materiálů.

Zdroje:

<http://interval.cz/clanky/kontrola-ceske-gramatiky-pro-ms-office-konec-korektoru-v-cechach/>

<http://www.scienceworld.cz/clovek/kontrolor-ceske-gramatiky-v-microsoft-office-1806/>