

# ZPRÁVA Z EXKURZE DO FIRMY LEXICAL COMPUTING

**Datum konání:** 13. 10. 2023

**Místo konání:** Lexical Computing, Botanická 554, Brno (FI MU)

**Název přednášky:** Exkurze do Lexical Computing CZ

**Přednášející:** Ondřej Matuška

**Počet účastníků:** 12

**Zpracoval:** Krňávek David

## O firmě

Firma Lexical Computing, která je na světové špičce v oblasti zpracování přirozeného jazyka, byla založena ve Velké Británii Adamem Kilgarriffem v roce 2003. Její česká pobočka sídlící v prostorách Fakulty informatiky Masarykovy univerzity v Brně se stala samostatnou firmou Lexical Computing CZ s.r.o. v roce 2011 a tato pobočka je dnes sídlem celé společnosti, ačkoli britská pobočka v Brightonu dle stránek firmy existuje dodnes.

Ředitelem Lexical Computing je Miloš Jakubíček. Ve firmě pracují výzkumní pracovníci zejména z Centra zpracování přirozeného jazyka Fakulty informatiky MU. Jelikož je firma strategickým partnerem FI MU, její tým se skládá převážně z absolventů této fakulty. Kromě nich zde pracují také absolventi programu Počítačová lingvistika na Filozofické fakultě MU.

Vlajkovým produktem firmy je korpusový manažer Sketch Engine, používaný vědci i studenty z celého světa. Kromě komerční verze existuje také verze s otevřeným zdrojovým kódem zvaná NoSketch Engine, v níž oproti Sketch Engine chybí některé funkce (např. nástroj pro analýzu chování slova *Word Sketch*). NoSketch Engine slouží také jako základ pro jiné korpusové manažery, např. KonText vytvořený v rámci Českého národního korpusu.

## O přednášejícím

Ondřej Matuška je marketingovým ředitelem firmy Lexical Computing. Má na starost prodej, marketingové aktivity a externí komunikaci, a to jak se zákazníci hledajícími informace o Sketch Engine, tak se stávajícími uživateli. Sám se představil jako „prodavač jazykových dat“.

## Průběh exkurze

Exkurze začala ve 13 hodin a konala se v sídle firmy v pátém poschodí CERIT Science Parku, přístavby Fakulty informatiky. Hned při příchodu mě překvapila poměrně neformální a uvolněná atmosféra, která panovala v prostoru firmy poskytující jazyková data zákazníkům z celého světa. Celé pracoviště se nachází v jedné velké místnosti rozdělené na dvě části. V části nacházející se blíže u vchodu do místnosti je umístěna výbava k promítání a půlkulatý stůl a tato část patrně slouží pro konání akcí podobných naší exkurzi. Ve druhé části, dále od vchodu, se nacházejí pracovní stoly s počítači. V době konání exkurze byla většina počítačů neobsazená.

V úvodu exkurze přednášející stručně představil firmu. Zmínil, že firma slaví v době konání exkurze 20 let, specializuje se na miliardové webové korpusy stahované na základě seznamu webových adres a nemá nic společného se samotnou Masarykovou univerzitou. Většina zákazníků Lexical Computing pochází ze zahraničí a přibližně 95 procent textů zpracovávaných firmou je v angličtině, českých jazykových dat je jen velmi malý podíl. Prvními zákazníky byla britská nakladatelství vydávající slovníky. Zákazníkem firmy byl dříve i Ústav pro jazyk český Akademie věd ČR, avšak to je již minulostí. Mnoho zahraničních jazykových ústavů včetně slovenského Jazykovedného ústavu Ludovíta Štúra SAV jsou však jejími zákazníky dodnes. Další zákazníci přednášející zmínil při jmenování příkladů užití jazykových dat, jež firma poskytuje. Společnost Amazon využívá řešení Lexical Computing pro analýzu recenzí produktu za účelem jeho zlepšení. Častým využitím jazykových dat je také extrakce terminologie za účelem sjednocení termínů v překladech různých překladatelů. To se týká např. uživatelských manuálů k výrobkům nebo právních a úředních dokumentů (tuto funkci integrovanou ve Sketch Engine používají zákazníci jako Bosch, Liebherr nebo Evropská komise). Jako další užití lze jmenovat také vyhledávání v mapách nebo GDEX (good dictionary examples, dobré slovníkové příklady) použitelné při tvorbě slovníku nebo výuce jazyka.

Problematickou záležitostí v oboru zpracování přirozeného jazyka, zejména v takto mezinárodním měřítku, jsou autorská práva. Na užití cizích textů za tímto účelem legislativa různých zemí pohlíží odlišným způsobem. Činnost společnosti Lexical

Computing i dalších společností zpracovávajících přirozený jazyk je obtížně právně definovatelná a obvykle se řadí do šedé zóny.

## Závěr

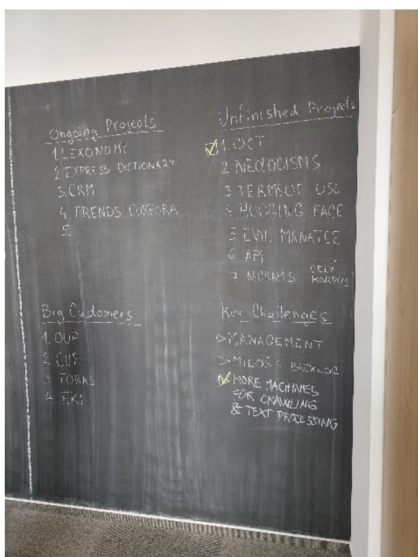
Z exkurze jsem si, stejně jako ostatní účastníci, odnesl mnoho zajímavých faktů ze „zákulisí“ jedné z velmi významných firem v oblasti zpracování přirozeného jazyka a dozvěděl jsem se o jejich projektech a zákaznících. V sídle takto vlivné firmy mě překvapila neformální atmosféra, která v prostorách panovala.



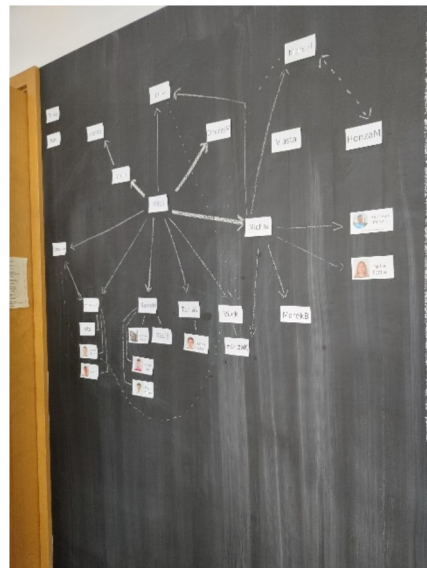
Lexical Computing je držitelem certifikátů Taylorcox



Prezentace projektů Lexical Computing



Tabule se seznamem projektů



Hierarchie zaměstnanců