

## I. IDENTIFIKAČNÍ ÚDAJE

<b>Název práce:</b>	<b>Generování přirozeného jazyka ze strukturovaných dat</b>
<b>Jméno autora:</b>	<b>Martin Matulík</b>
<b>Typ práce:</b>	diplomová
<b>Fakulta/ústav:</b>	Fakulta elektrotechnická (FEL)
<b>Katedra/ústav:</b>	Katedra počítačů
<b>Oponent práce:</b>	Jan Pichl
<b>Pracoviště oponenta práce:</b>	Katedra kybernetiky

## II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

<b>Zadání</b>	<b>náročnější</b>
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Generování přirozeného jazyka patří k jednomu z nejnáročnějších úkolů z oblasti zpracování přirozeného jazyka. Základní požadavkem je gramatická správnost věty. Často jsou na výstup kladeny další požadavky typu pořadí informací, shlukování podobných informací atd. Z tohoto důvodu hodnotí zadání jako náročnější.	

<b>Splnění zadání</b>	<b>splněno</b>
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Splněno bez výhrad.	

<b>Zvolený postup řešení</b>	<b>správný</b>
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Student zvolil správný postup řešení, který je inspirován současnými technikami "sequence-to-sequence" generováním textu, kdy na základě vstupní sekvence a doposud vygenerované části výstupní sekvence je určena pravděpodobnost následujícího slova.	

<b>Odborná úroveň</b>	<b>B - velmi dobře</b>
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Práce je na dobré odborné úrovni. Student prováděl experimenty na modelu, který je inspirován úspěšnými publikacemi. Předpokládám, že výběr modelu a parametrů by realizován s ohledem na velice dlouhou dobu trénování těchto NLG modelů, přesto bych ocenil v práci delší diskusi na toto téma.	

<b>Formální a jazyková úroveň, rozsah práce</b>	<b>B - velmi dobře</b>
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Po jazykové stránce je práce na výborné úrovni. Vytкнуł bych pouze drobnosti, např. použití špatných úvodních uvozovek, na str. 18 nekonzistentní uvádění citací před/za tečkou věty. V závěru práce student uvádí tabulky se dvěma příklady, které by bylo vhodnější umístit do předcházející sekce a pak se na ně pouze odkazovat.	

<b>Výběr zdrojů, korektnost citací</b>	<b>A - výborně</b>
<i>Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.</i>	
Uvedené zdroje jsou v pořádku. Student uvádí citace všude, kde je potřeba.	

<b>Další komentáře a hodnocení</b>
------------------------------------

*Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.*

### **III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE**

*Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.*

Student se zabývá velice těžkým úkolem z oblasti zpracování přirozeného jazyka. Základní myšlenkou je vytvořit větu v přirozeném jazyce s použitím vstupních strukturovaných dat. Student nejprve v teoretické části popisuje jazykové modely a přístupy ke generování přirozeného jazyka. Dále navrhuje architekturu modelu inspirovanou existujícím řešením. Student experimentuje s nastavením několika hyperparametrů, avšak mnoho z nich nechává na původních hodnotách použitých v odkazovaném článku. To se dá vysvětlit velice dlouhou dobou tréninku, která v oblasti generování přirozeného jazyka není nic neobvyklého, avšak toto zdůvodnění v práci chybí.

Otázka na studenta:

Na jakých datech byste model trénoval, pokud byste ho chtěl použít pro generování vět (dialogů) z existujících znalostních databází (Freebase, Wikidata)?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **B - velmi dobře.**

Datum: 13.6.2018

Podpis: