



## DRUŽBA / MEDIJI

# Novinarjem v pomoč umetna inteligenca

✂ **Uroš Škerl Kramberger**  
**V Ljubljani poteka srečanje partnerjev evropskega projekta Embeddia, v katerem bodo s pomočjo umetne inteligence analizirali delovne medijev. Namen je ustvariti računalniška orodja, ki bodo novinarjem pomagala hitreje pisati novice o dogodkih, ki se zgodijo v manjših evropskih članicah.**

Če bi zaradi pritiska protestnikov na Madžarskem padla vlada, bi bilo to nemudoma objavljeno na katerem od tamkajšnjih neodvisnih medijskih portalov. V Sloveniji pa bi bila novica – četudi je Madžarska sosednja država in četudi se za premierja Viktorja Orbana zanima velik del javnosti – objavljena šele z zamikom. Le redko kateri slovenski novinar govori madžarsko in slovenski novičarski servisi bi prve interpretacije ter izjave najverjetneje povzemali po angleških tiskovnih agencijah ali globalnih servisih, kot je BBC. Takšna novica bi bila nekoliko zakasnela, zaradi vmesnega prevajanja v »tretji« jezik, angleščino, pa morebiti tudi ne dovolj natančna.

## Malim jezikom vse, kar imajo veliki

Novinarji, ki ustvarjajo novice v večjih evropskih jezikih, so v prednosti, saj jim takšnih »ovinkov« ni treba uporabljati. Poleg tega veliki medijski servisi že uporabljajo računalniško podprte jezikovne tehnologije, ki s pomočjo metod umetne inteligence analizirajo izvorna besedila in novinarjem pomagajo hitro prevajati ter oblikovati sporočila. Za manjše jezike, ki jih je v

Evropi veliko, pa so tovrstni pripomočki še v fazi razvoja. Zato je Evropska unija podprla projekt s področja umetne inteligence Embeddia: Medjezikovne vektorske vložitve za manj zastopane jezike v evropskih medijih (Cross-Lingual Embeddings for Less-Represented Languages in European News Media). Vodenje projekta je prevzel odsek za tehnologije znanja na **Inštitutu Jožefa Stefana** (IJS), zanj je namenjenih tri milijone evrov, potekal pa bo tri leta. V njem bodo poleg slovenskih sodelovali partnerji s Finskega, Hrvaškega, iz Estonije ter Francije in Velike Britanije. Uvodno srečanje projektnih partnerjev trenutno poteka v Ljubljani. Ob raziskovalnih institucijah bodo vanj vključena tudi štiri evropska podjetja, med njimi na Hrvaškem registrirana podružnica avstrijskega medijskega podjetja Styria (to je tudi solastnik *Dnevnik*).

Že razvita računalniška orodja, ki delujejo v večjih jezikih, kot sta francoščina in angleščina, bodo služila kot osnova za razvoj orodij za manjše jezike. »Razvili bomo metode za poravnave besed med jeziki, ki bodo omogočile hiter prenos orodij z večjih na manjše jezike,« pravi koordinatorka projekta dr. **Senja Pollak**, strokovnjakinja za jezikovne tehnologije, ki raziskuje na IJS in na Univerzi v Edinburgu.

»Odgovarjali bomo na sodobne izzive medijske industrije, ukvarjali se bomo z medjezikovno analizo novic, analizo spletnih komentarjev in z razvojem orodij za pomoč pri avtomatiziranem tvorjenju novic. Če želimo na primer novice hitro posredovati v različnih jezikih, so za to ključna orodja za pomoč pri avtomatiziranem tvorjenju novic iz vhodnih podatkov. V okviru projekta bomo analizirali potrebe pri produkciji novic in kreativni uporabi jezika, na primer pri tvorjenju

naslovov. Naš namen ni, da bi izdelali popolnoma avtomatska orodja, temveč orodja, ki bi novinarjem olajšala delo. Pomemben pa je tudi poudarek na etičnem novinarstvu in raziskovanju,« pravi Senja Pollak.

## Z umetno inteligenco nad sovražni govor

»Urednikom bomo pomagali hitreje zaznavati žaljivke, vulgarne izraze in izraze, ki spadajo na področje sovražnega govora,« pravi profesor **Marko Milosavljevič** z ljubljanske fakultete za družbene vede (FDV), ki bo tudi sodelovala pri projektu. »Žaljivi izrazi v angleščini, denimo besedi f\*\*\* in nigg\*\*, so že dobro obdelani. S pomočjo metod umetne inteligence jih lahko uredniki angleških spletnih medijev naglo zaznajo in jih po potrebi odstranijo. Pri tem ne gre za cenzuro vsebine, ampak za odstranjevanje nestrpnih, žaljivih izrazov in klevtic,« pripoveduje Milosavljevič. »S predlaganimi tehnologijami pa lahko program za prepoznavanje sovražnega govora, ki je bil naučen ob velikih količinah podatkov, iz angleščine z manjšimi prilagoditvami hitro prenesemo v slovenščino,« pojasnjuje profesor **Marko Robnik - Šikonja** s Fakultete za računalništvo in informatiko Univerze v Ljubljani, ki je tehnični koordinatorka projekta.

Projekt Embeddia bo usmerjen tudi na druga področja – med drugim v avtomatsko analiziranje podatkov in računalniško generiranje besedil, ki ne zahtevajo poglobljenega novinarskega dela, temveč zgolj prepisovanje ali povzemanje sporočil za javnost, dodaja Milosavljevič. Z razvojem tovrstnih računalniških orodij bo novinarjem ostalo več časa za pravo preiskovalno oziroma terensko delo, kakršnega računalnik seveda ne zmore. ✕

**Projekt Embeddia bo med drugim usmerjen v avtomatsko analiziranje podatkov in računalniško generiranje besedil, ki ne zahtevajo poglobljenega novinarskega dela. Z razvojem tovrstnih računalniških orodij bo novinarjem ostalo več časa za pravo preiskovalno oziroma terensko delo, kakršnega računalnik seveda ne zmore.**