

Příloha č. 1 – Závazné parametry řešení projektu

1. Název projektu

Název projektu

Webový pravopisný, gramatický a typografický korektor pro český jazyk

2. Datum zahájení a ukončení projektu

Datum zahájení a ukončení projektu

02/2019 - 05/2022

3. Předmět řešení návrhu projektu

Předmět řešení návrhu projektu

Cílem projektu je připravit komplexní automatický korektor textů psaných českým jazykem, který bude schopen pokrývat nejen kontrolu překlepů, ale také pravopisných, gramatických a typografických chyb. Bude předložen v podobě veřejně dostupného webového rozhraní, do kterého bude možné psát text jako ve standardním editoru nebo nahrát již hotový textový soubor, a ten předložit k opravě. V případě zájmu se uživatel bude moci seznámit s podrobnějším vysvětlením, proč byla daná sekvence textu označena jako chybná. V této oblasti počítáme s propojením s výkladovou částí Internetové jazykové příručky (IJP), ve které je přehledně vysvětlena řada pravopisných, gramatických a typografických chyb. Opravy chyb typografických se budou opírat částečně o IJP, a především o ČSN 01 6910.

4. Cíl projektu (účel podpory)

Cíl projektu (účel podpory)

Projekt bude řešen ve spolupráci tří akademických pracovišť, MU, ÚJČ a UK, a interního aplikačního garanta, společnosti Seznam.cz.

Spolupráce je založena na kooperaci odborníků a sdílení know-how, jazykových dat, softwaru, hardwaru a některých již existujících řešení dílčích problémů ve vývoji jazykového korektoru.

V současné době je možné tuto spolupráci realizovat díky vysokému stupni vývoje automatických nástrojů pro zpracování češtiny, dostatku elektronických strukturovaných jazykových dat, existujícím dílčím výsledkům v oblasti automatických oprav českých textů a také díky možnosti spojení jak zkušených odborníků, tak mladých perspektivních doktorandů počítačové lingvistiky v jednom projektu.

Inovativnost výsledku lze shrnout do několika bodů. Jazykový korektor:

- bude veřejně dostupný jako webové rozhraní,
- za jeho kvalitu budou ručit relevantní akademická pracoviště,
- kromě překlepů bude zaměřen i na nejméně frekventovanější pravopisné, gramatické a (omezeně) stylistické chyby,
- bude schopen pokrýt velkou část chyb typografických v podobě hromadných i volných oprav,
- bude i po skončení projektu soustavně udržován a inovován (meze automatických oprav lze stále posunovat o malé kroky vpřed).

Práce korektoru bude založena na morfologickém označování textu (každému slovu je přidělen kód, který nese informace o jeho tvaru), desambiguaci (zjednodušení) přidělených značek (mnoho tvarů slov je homonymních a při analýze obdrží více značek) a syntaktické analýze, do které budou implementována pravidla pro rozpoznání chyb.

MU je zastoupena Ústavem českého jazyka (ÚČJ) FF a Centrem zpracování přirozeného jazyka (CZPJ) FI.

Do projektu vkládá:

- odborné znalosti češtiny, počítačové lingvistiky a počítačového zpracování přirozeného jazyka,
- jazyková data v podobě sady valenčních rámců,
- software pro zpracování české morfologie a syntaxe (ajka, desamb, SET),
- hardware (provoz výsledného korektoru bude zajišťovat server na FI MU),
- dílčí řešení v podobě pravidel pro opravu interpunkce, vybraných gramatických chyb a pro hromadné opravy chyb typografických (magisterské práce a disertace studentů počítačové lingvistiky).

ÚJČ AV ČR do projektu poskytuje:

- zkušenosti v oblasti kodifikace češtiny, jazykového poradenství, vývoje a provozu IJP,
- slovníková data – hesláře a tvary slov z IJP a část neologismů z databáze Neomať,
- textová vysvětlení jazykových jevů z výkladové části IJP a možnost propojení příručky s

korektorem,

- detailní znalosti pravidel typografie (v projektu jsou zapojeni autoři ČSN 01 6910),
- zajištění publicity pro výsledný korektor.

UK zastoupená Ústavem teoretické a počítačové lingvistiky FF poskytuje:

- znalosti z oblasti formálního (pravidlového) popisu češtiny,
- zkušenosti s tvorbou korektoru pravopisu a gramatiky v MS Word,
- zkušenosti s přípravou pravidel pro automatickou desambiguaci,
- software pro morfologickou analýzu a desambiguaci (MorphoDiTa).

Aplikační garant Seznam.cz do projektu vkládá:

- zkušenosti z oblasti automatické syntaktické analýzy s využitím slovesné valence,
- zkušenosti z vývoje interního korektoru překlepů,
- rozsáhlá jazyková data k testování verzí korektoru.

Práce na jazykovém korektoru nebudou probíhat izolovaně na jednotlivých pracovištích, členové řešitelského týmu budou propojeni díky online komunikaci v dílčích týmech napříč zapojenými subjekty. V harmonogramu je navržen souběh činností tak, aby byla zajištěna co nejvyšší efektivita celého vývoje.

T A

Č R

ZÁVAZNÉ PARAMETRY ŘEŠENÍ PROJEKTU

Číslo projektu: **TL02000146**

5. Osoby

Klíčová osoba řešitelského týmu

Role Řešitel			
Tituly před jménem ██████	Jméno ████	Příjmení ██████████	Tituly za jménem ████

6. Harmonogram a výstupy/výsledky projektu

Pracovní verze

7. Identifikační údaje účastníků

Hlavní příjemce - [P] Masarykova univerzita

IČ 00216224	DIČ CZ00216224	Obchodní jméno Masarykova univerzita
Organizační jednotka Filozofická fakulta		Kód organizační jednotky 14210
Právní forma VVS - Veřejná nebo státní vysoká škola (zákon č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů)		
Rodné číslo	Typ organizace VO - Výzkumná organizace	Typ VO VVS - veřejná vysoká škola

Další účastník - [D] Ústav pro jazyk český AV ČR, v.v.i.

IČ 68378092	DIČ CZ68378092	Obchodní jméno Ústav pro jazyk český AV ČR, v.v.i.
Organizační jednotka		Kód organizační jednotky
Právní forma VVI - Veřejná výzkumná instituce (zákon č. 341/2005 Sb., o veřejných výzkumných institucích)		
Rodné číslo	Typ organizace VO - Výzkumná organizace	Typ VO AV ČR - Akademie věd ČR

Další účastník - [D] Univerzita Karlova

iČ 00216208	DIČ CZ00216208	Obchodní jméno Univerzita Karlova
Organizační jednotka Filozofická fakulta		Kód organizační jednotky 11210
Právní forma VVS - Veřejná nebo státní vysoká škola (zákon č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů)		
Rodné číslo	Typ organizace VO - Výzkumná organizace	Typ VO VVS - veřejná vysoká škola

Další účastník - [D] Seznam.cz, a.s.

iČ 26168685	DIČ CZ26168685	Obchodní jméno Seznam.cz, a.s.
Organizační jednotka		Kód organizační jednotky
Právní forma POO - Právnícká osoba zapsaná v obchodním rejstříku (zákon č. 304/2013 Sb., o veřejných rejstřících právnických a fyzických osob)		
Rodné číslo	Typ organizace VP - Velký podnik	

Pravidla pro publicitu projektů podpořených z prostředků TA ČR

1. ÚVOD

Pravidla pro publicitu (dále jen Pravidla) stanovují povinné náležitosti v oblasti zajištění propagace projektů a jejich výsledků a výstupů podpořených Technologickou agenturou České republiky (dále jen TA ČR) u příjemců podpory. Rozsah publicity závisí na typu jednotlivých výsledků a výstupů realizovaných při řešení konkrétního projektu.

Každý příjemce podpory a další účastníci podílející se na řešení projektu realizovaného s podporou z prostředků TA ČR musí v případě, že prezentují informace o jimi řešeném projektu s podporou TA ČR v hromadných sdělovacích prostředcích, informovat o tom, že projekt byl realizován za finanční spoluúčasti TA ČR, a to na všech propagačních materiálech i ve všech typech médií, které se k projektu či jeho výsledkům a výstupům vztahují.

2. VYSVĚTLENÍ POJMŮ

Logotyp je specifický grafický symbol, který identifikuje firmu, obchodní značku nebo produkt doplněný o text.

Impresum, zvané též autorská tiráž nebo vydavatelský záznam, je soubor povinných informací o knize, který je umístěn na stránce následující za hlavním titulem knihy. Obsahuje povinně (v ČR) copyrightové záznamy o knize (kdo je majitelem autorských práv v textu, ilustracím apod.) a ISBN.

Tiráž obsahuje údaje o knize, a dá se nazvat jako tzv. „rodný list knihy“. Jedná se o soupis informací obsahujících technické a vydavatelské informace o knize. Tiráž je zpravidla umístěována na poslední stranu knihy. Nesprávně bývá také umístěována na začátku knihy místo impresum. Tiráž obsahuje informace povinné (podle zákona č. 37/1995 Sb., o neperiodických publikacích) a nepovinné.

Copyright označuje fakt, že dané dílo je chráněným autorským zákonem. Obvykle se označuje symbolem © následovaným rokem a jménem autora. Tento symbol má však v ČR i ve většině jiných zemí pouze ryze informativní význam – dílo je plně chráněno i tehdy, není-li to na něm nijak uvedeno.

Plakát je list papíru, který zprostředkovává komunikaci mezi výrobcem, prodávajícím a zákazníkem. Plakát s sebou nese jak textovou, tak také vizuální informaci. Brožura je jednoúčelová brožovaná neperiodická tiskovina, obvykle speciální povahy a účelu.

Panel/Banner je deska (často z PVC) pro připevnění exponátů nebo výzdoby na výstavách a expozicích.

3. PRAVIDLA PRO PUBLICITU

Příjemce i další účastník je povinen informovat veřejnost o tom, že projekt, resp. Jeho výstupy a výsledky, byl spolufinancován z prostředků TA ČR. To je možné zajistit slovním vyjádřením informace, že projekt byl spolufinancován z prostředků TA ČR (přesné vzory formulací jsou uvedeny v části 3 těchto Pravidel) anebo použitím logotypu (loga) TA ČR.

Pravidla publicity se vztahují na všechny typy výsledků a výstupů podporovaných ze všech programů TA ČR. Součástí publicity musí být logotyp TA ČR či text (dle typu výsledku), a to na každém výstupu spojeném s propagací projektu či jeho výsledků a výstupů.

Informace o spolufinancování by se tedy měla objevit především na dokumentech a dalších materiálech určených veřejnosti, zejména na prezentacích a dalších nosičích prezentovaných nebo šířených na konferencích či seminářích (např. letáky, pozvánky, informační brožury), v odborných člancích či publikacích.

Informace o spolufinancování by měla být uvedena také na přístrojích pořízených z prostředků projektu, pokud to jejich charakter a rozměry umožňují. U přístrojů či jiného materiálového vybavení doporučujeme informovat prostřednictvím loga umístěného přímo na přístroji (např. formou samolepky).

Informace nemusí být uvedena na materiálech či na dokumentech, na nichž to vzhledem k jejich povaze není možné (např. patentová přihláška).

Náklady spojené se zajištěním publicity nejsou způsobilým nákladem projektů TA ČR. Příjemce podpory i další účastníci projektu jsou povinni řídit se těmito Pravidly a uchovávat pro potřeby kontroly doklady o tom, že byly dodrženy podmínky uvedené v těchto Pravidlech (např. brožury, letáky, plakáty, fotodokumentaci, samolepky aj.), a to po dobu řešení projektu a po dobu 10 let od jeho ukončení.

4. NÁLEŽITOSTI INFORMACE O PODPOŘE A PRAVIDLA UŽITÍ LOGOTYPU TA ČR

4.1 Slovní informace

Pro informování o spolufinancování projektu či jeho výsledku z prostředků TA ČR je možné použít následující slovní spojení:

„Projekt (číslo/název projektu) je/byl řešen s finanční podporou TA ČR.“

„Tato metodika (doplnit konkrétní výsledek) byla vytvořena s finanční podporou TA ČR“.

„Tento software (doplnit konkrétní výsledek) byl vytvořen s finanční podporou TA ČR“.

„Tyto specializované mapy (doplnit konkrétní výsledek) byly vytvořeny s finanční podporou TA ČR“.

Nejvhodnější spojení (upravené samozřejmě dle typu výsledku) zvolí příjemce dle svého uvážení.

4.2 Logotyp

Pro potřeby správného používání logotypu TA ČR slouží Grafický manuál pro příjemce, který je ke stažení na http://www.tacr.cz/dokums_raw/3.11.2._Grafick%C3%BD_manu%C3%A1l_prijemci01.pdf, určující pravidla, způsob použití a zveřejnění logotypu TA ČR. Logotyp TA ČR je důležitým nástrojem komunikace a musí být nedílnou součástí všech propagačních a informačních výsledků/výstupů z projektů spolufinancovaných z prostředků TA ČR. Logotyp TA ČR je ke stažení na <http://www.tacr.cz/tiskove-centrum/ke-stazeni/>

Rozměr logotypu zvolí příjemce či další účastník individuálně, pro každý projekt/výstup samostatně, tak aby byla zajištěna jeho dostatečná viditelnost a čitelnost. Na předměty je možné logotyp umístit např. pomocí samolepky.

5. TYPY VÝSLEDKŮ

V následující části znění Pravidel jsou uvedeny možné způsoby informování o podpoře z prostředků TA ČR na jednotlivých typech výsledků podporovaných v programech TA ČR. Nejsou zde uvedeny dílčí výstupy projektu, jako zpracovaná rešerše, protokoly z laboratorních zkoušek či protokol o zkouškách či zpracované výsledky šetření. V závěru této kapitoly jsou uvedeny i další (nepodporované) výsledky dle RIV, které však mohou při řešení projektů TA ČR vzniknout.

P – patent

Pokud je výsledkem projektu udělený patent, není z pochopitelných důvodů možné uvádět logo TA ČR přímo na patentové listině či žádosti o udělení patentu. Nicméně, informace o obdržené podpoře by měla být uvedena v materiálech, ve kterých se o předmětu patentu hovoří, tedy např. na propagačních materiálech jako jsou brožury, letáky či plakáty. Informace by měla být zmíněna také v článcích či prezentacích týkajících se předmětu patentu. Informace musí být viditelná a text čitelný.

Z – poloprovoz, ověřená technologie, F – užitný vzor, průmyslový vzor, G – prototyp, funkční vzorek

Pokud je výsledkem projektu poloprovoz, ověřená technologie, užitný vzor, průmyslový vzor, prototyp či funkční vzorek, uvádí se informace o obdržené podpoře vč. logotypu TA ČR na materiálech, ve kterých se o těchto výsledcích hovoří, například na plakátech. Informace by měla být zmíněna také v článcích nebo prezentacích týkajících se poloprovozu či ověřené technologie. Informace musí být viditelná a text čitelný.

N – certifikované metodiky a postupy, vč. specializovaných map s odborným obsahem

Pokud jsou výsledkem projektu certifikované metodiky a postupy, vč. specializovaných map s odborným obsahem, uvádí se viditelná informace o tom, že projekt byl realizován za finanční spoluúčasti TA ČR na druhé straně titulní strany materiálu v tištěné podobě (viz pojem impresum).

T A Č R

Pokud budou výsledkem projektu mapy v elektronické podobě, využijí se pravidla uvedená u výsledku typu R – software. Informace musí být viditelná a text čitelný.

H – výsledky promítnuté do právních předpisů a norem, do směrnic a předpisů nelegislativní povahy

Pokud jsou výsledkem projektu právní předpisy a normy nebo směrnice a předpisy nelegislativní povahy, uvádí se viditelná informace o tom, že projekt byl realizován za finanční spoluúčasti TA ČR na druhé straně titulní strany materiálu v tištěné podobě (viz pojem impresum), který je předán příslušnému orgánu státní správy k promítnutí výsledků do právních předpisů. Informace musí být viditelná a text čitelný.

R – software

Pokud je výsledkem projektu software, musí být informace o podpoře (nejlépe logotyp TA ČR) uvedena vždy při spuštění příslušného programu. Logotyp také může být umístěn na samotném nosiči (např. CD) či na jeho obalu. Informace musí být viditelná a případný text čitelný.

V – výzkumná zpráva

Vzhledem k tomu, že se u tohoto výsledku musí vždy jednat o výzkumnou zprávu s utajovaným obsahem, která je ze své povahy určena pouze omezenému okruhu adresátů, není umístění informace o spolufinancování povinné.

Článek

Informace o podpoře TA ČR musí být uvedena v rámci textu článku, na konci článku či v odkazech, v závislosti na koncepci článku.

Publikace/sborník

Pokud jsou výsledkem projektu publikace či sborníky, uvádí se viditelná informace o tom, že projekt byl realizován za finanční spoluúčasti TA ČR, na druhé straně titulní strany (viz pojem impresum).

Tisková zpráva

Informace o podpoře TA ČR musí být uvedena v rámci textu zprávy, ideálně v prvním odstavci. V případě, že je projekt či výsledky a výstupy podpořený výhradně z prostředků TA ČR, musí být na tiskové zprávě umístěno také logotyp TA ČR.

6. TECHNICKÉ PROSTŘEDKY PROPAGACE

K propagaci projektu, jeho výsledků a výstupů je možné použít následující propagační prostředky. U každého z nich uvádíme formální náležitosti, které takový prostředek musí splňovat.

Prezentace (PowerPoint a jiné)

Pokud se prezentace týká pouze projektu či jeho výstupu, musí být informace o spolufinancování z prostředků TA ČR uvedena na titulní straně prezentace. Současně by v prezentaci mělo být uvedeno i číslo a název projektu. Pokud je o projektu či jeho výsledcích v prezentaci pouze zmínka (a prezentace jako celek je zaměřená na odlišné téma), musí být informace o tom, že tento projekt byl realizován za podpory/z prostředků TA ČR, uvedena na snímku, na kterém se o projektu či jeho výstupech hovoří poprvé.

Internetové stránky projektu nebo informace o projektu na stránkách příjemce

Na internetových stránkách projektu či u informací o projektu na stránkách příjemce musí být na úvodní straně či v úvodu informace o projektu uvedeno sdělení o spolufinancování projektu TA ČR vč. logotypu. Informace musí být uvedena také na webových stránkách, na kterých je prezentován výsledek projektu. Stránky také musí obsahovat odkaz na webové stránky poskytovatele podpory, například formou aktivního odkazu ze slovního spojení TA ČR či logotypu.

Brožura

Brožura musí obsahovat informaci o spolufinancování TA ČR vč. logotypu. Informace musí být umístěna viditelně buď na titulní straně brožury, nebo na druhé straně její titulní strany (viz pojem impresum). V případě zásadního narušení designu či estetického charakteru titulní strany brožury je možné informace o spolufinancování umístit na poslední stranu brožury.

Leták

V případě letáku se použije doporučení shodné jako pro brožuru jen s tím rozdílem, že informace o spolufinancování a logotyp by měly být vždy na titulní straně letáku a vždy ve viditelné velikosti – minimální velikost je stanovena grafickým manuálem.

Plakát

Velikost plakátu musí dosahovat minimálně rozměru A4, přičemž z celkové plochy plakátu musí být vždy minimálně 5 % určeno pro informaci o spolufinancování projektu z prostředků TA ČR vč. logotypu. Je-li plakát použit jako prostředek publicity pro projekt, jehož výsledkem je poloprovoz, je plakát umístěn v tomto prostoru. Je-li plakát použit jako prostředek publicity pro projekt, jehož výsledkem je ověřená technologie, užitečný vzor či patent, je plakát umístěn v blízkosti této technologie, případně viditelně na odpovídajícím místě v provozovně příjemce, ve které je tento podpořený majetek využíván.

T A

Č R

Konference, semináře, veletrhy, výstavy (dále jen akce)

Pokud je součástí akce stánek, musí být opatřen logotypem TA ČR event. plakáty. Pokud jsou součástí akce panely/bannery, musejí být umístěny na dostatečně viditelném místě a musejí obsahovat plochu vyhrazenou pro informaci o finanční podpoře TA ČR vč. logotypu. Pokud jsou v souvislosti s danou akcí vytvořeny také tištěné dokumenty (např. prezentace, brožury), platí doporučení uvedená pro jednotlivé typy dokumentů.

Schválil: předsednictvo TA ČR

Dne: 12. února 2015