



Grantová agentura České republiky

Část A

Návrh standardního grantového projektu (dále návrh projektu)

Datum vytištění přihlášky: 21.04.2009	Číslo panelu(ů) P406	Registrační číslo: P406/10/0434
---	--------------------------------	---

UCHAZEČ a NAVRHOVATEL

Uchazeč:	Univerzita Karlova v Praze			IČ:	00216208
Sídlo:	Ovocný trh 5, Praha 1, 11636				
Navrhovatel:	doc. RNDr. Vladimír Petkevič, CSc.				
Pracoviště navrhovatele:	Filozofická fakulta, Praha 1, nám. Jana Palacha 2				
Rodné číslo:	Telefon:	Fax:	E-mail:		
5403023274	+420 221 619 753	+420 221 619 753	vladimir.petkevic@ff.cuni.cz		

Název projektu česky:	
Syntaktická anotace českých korpusů	
Název projektu anglicky:	
Syntactic annotation of Czech corpora	
Klíčová slova česky:	korpusová lingvistika; značkování korpusů; mělká syntaktická analýza
Klíčová slova anglicky:	corpus linguistics; tagging; shallow parsing
Datum zahájení:	01/01/2010
Doba řešení (roky):	3
Zařazení do číselníku CEP:	AI,Jazykověda

Hodnocení grantových projektů GA ČR ukončených v posledních třech letech, u kterých byl navrhovatel řešitelem

Registrační číslo:		
Registrační číslo:		
Registrační číslo:		
Registrační číslo:		

[illegible]

Kopie oprávnění k činnosti tvořící součást řešení grantového projektu ve smyslu Zadávací dokumentace čl. 4.2.1.

NENÍ PŘEDMĚTEM PROJEKTU

Uchazeč a navrhovatel svými podpisy na návrhu projektu stvrzují, že:

- se před podpisem návrhu projektu seznámili s obsahem Zadávací dokumentace pro veřejnou soutěž ve výzkumu a vývoji na podporu grantových projektů základního výzkumu na rok 2010 - standardní grantové projekty - (dále Zadávací dokumentace nebo ZD) a Grantového systému GA ČR 2009 (dále GS) a zavazují se dodržovat jejich ustanovení;
- návrh projektu byl vypracován v souladu se Zadávací dokumentací a GS;
- všechny údaje uvedené v podepsaném výtisku návrhu projektu jsou pravdivé, úplné a nezkreslené a jsou totožné s údaji v elektronické verzi návrhu projektu podané pomocí aplikace;
- všichni spoluuchazeči, spolunavrhovatelé a spolupracovníci uvedení v návrhu projektu byli seznámeni s věcným obsahem návrhu projektu, jakož i s finančními požadavky uvedenými v návrhu projektu a dále i s obsahem ZD a GS;
- před podáním návrhu projektu zajistili souhlas výše uvedených osob s účastí na řešení grantového projektu uvedeného v návrhu projektu za v něm specifikovaných podmínek;
- souhlasí, aby údaje uvedené v tomto návrhu projektu byly použity pro vnitřní informační systém GA ČR a dále souhlasí s uveřejněním údajů v rozsahu stanoveném v ZD a GS.

.....
Podpis navrhovatele
(Petkevič)

Podpis statutárního orgánu a razítko uchazeče
(Michal Stehlík, děkan)

.....
Datum


Grantová agentura České republiky
Část Abstrakt

Navrhovatel:	Registrační číslo :
doc. RNDr. Vladimír Petkevič, CSc.	P406/10/0434
Název projektu:	
Syntaktická anotace českých korpusů	

Abstrakt (max. 1100 znaků) - česky

Projekt si klade za cíl vytvořit sadu nástrojů, s jejichž pomocí bude možné syntakticky anotovat korpusy současné češtiny. V současné době existují nástroje na morfologické značkování, které již byly využity pro označkování Českého národního korpusu. Existuje i několik systémů syntaktické analýzy češtiny, které jsou však příliš teoreticky zatížené a jejich výsledky nejsou proto vhodné pro značkování korpusů určených široké odborné i laické veřejnosti. Navíc jejich úspěšnost není dost vysoká na to, aby umožňovaly plně automatické zpracování textů. Předkládaný projekt počítá s využitím všech zmíněných existujících nástrojů a předpokládá nižší procento chyb po převodu výsledků do teoreticky neutrálnějšího formátu (odpovídajícímu šmilauerovské teorii vyučované na školách). Dále předpokládá vytvoření sady nástrojů pro opravování chyb a úplné syntaktické označkování pokusného korpusu o velikosti cca 1 mil. slov.

Cíle projektu (max. 300 znaků) - česky. (Tento text bude v případě udělení grantu uveden ve smlouvě o řešení projektu.)

Návrh značek pro syntaktickou anotaci synchronních českých korpusů, procedury pro převod výsledků syntaktické analýzy použité pro PDT do navržených značek, nástroje pro vyhledávání a opravování chyb v syntaktické anotaci, označkování pokusného korpusu o velikosti cca 1 mil. slov.

Abstrakt (max. 1100 znaků) - anglicky

The project's aim is to create a set of tools for syntactic annotation of Czech text corpora. Currently, there exist tools for morphological tagging, which were already used for tagging the Czech National Corpus. There are also several systems for syntactic analysis of Czech, but they are heavily dependent on the underlying syntactic theory, and thus their results are not suitable for the purpose of tagging corpora aimed at the general public. Beside it, their success rate is not high enough for fully automatic processing of texts. The proposed project will exploit all of the abovementioned tools. After converting the results to a format that will be syntactically more neutral, we expect a lower error ratio. Furthermore, we will also create a set of tools for error correction and then we will annotate an experimental corpus of the size ca 1 million words.


Grantová agentura České republiky
Část B-sumy

Uchazeč:
Navrhovatel:

Univerzita Karlova v Praze - Filozofická fakulta
doc. RNDr. Vladimír Petkevič, CSc.

Registrační číslo:
P406/10/0434

1) Celkové předpokládané uznané náklady na řešení projektu ze všech zdrojů financování na jednotlivé roky jeho řešení

(finanční údaje se uvádějí jako celočíselné hodnoty v **tis. Kč**)

	1. rok	2. rok	3. rok	Celkem
Náklady ze všech zdrojů financování	1455	1493	1575	4523

2) Celkové předpokládané uznané náklady na řešení projektu z jednotlivých zdrojů za celou dobu jeho řešení

(finanční údaje se uvádějí jako celočíselné hodnoty v **tis. Kč**)

Jednotlivé zdroje finančních prostředků na řešení projektu	tis. Kč
Celkové grantové prostředky požadované z GA ČR	4523
Podpora z jiných tuzemských veřejných zdrojů (z jiné kapitoly státního rozpočtu nebo rozpočtů územních samosprávných celků), pokud existuje	0
Podpora z ostatních veřejných zdrojů (nepatřících do státního rozpočtu nebo rozpočtů územních samosprávných celků), pokud existuje (veřejné zdroje v ČR i v zahraničí)	0
Podpora z neveřejných zdrojů (zahraniční zdroje, neveřejné tuzemské zdroje, vlastní neveřejné zdroje, pokud existuje)	0
Celkem	4523

3) Celkové náklady na řešení projektu požadované od GA ČR

(finanční údaje se uvádějí jako celočíselné hodnoty v **tis. Kč**)

	1. rok	2. rok	3. rok
Věcné náklady celkem	66	104	186
Investiční náklady celkem	0	0	0
Osobní náklady celkem	1389	1389	1389

Náklady na řešení projektu celkem	1455	1493	1575
-----------------------------------	------	------	------



Grantová agentura České republiky

Část B-rozpis

Finanční prostředky požadované od GA ČR pro uchazeče

(finanční údaje se uvádějí jako celočíselné hodnoty v tis. Kč)

Uchazeč:

Navrhovatel:

Univerzita Karlova v Praze - Filozofická fakulta

doc. RNDr. Vladimír Petkevič, CSc.

Registrační číslo:

P406/10/0434

		Požadované prostředky od GA ČR		
		1. rok	2. rok	3. rok
Věcné náklady 1)				
Materiální náklady		45	45	45
Cestovní náklady		0	30	60
Náklady na ostatní služby a nemateriální náklady		10	10	50
Doplňkové (režijní) náklady		11	19	31
Věcné náklady celkem		66	104	186
<u>Investiční náklady (na pořízení dlouhodobého hmotného a nehmotného majetku)</u>	Celková pořizovací cena			
2)				
		0	0	0
		0	0	0
		0	0	0
		0	0	0
		0	0	0
Investiční náklady celkem	0	0	0	0
<u>Osobní náklady (podrobný rozpis v části B-osobní náklady) 3)</u>	Počet			
Mzdy nebo odměny navrhovatele a spolupracovníků		6	890	890
Mzdy nebo odměny technických pracovníků			55	55
Ostatní osobní náklady (celkem)			80	80
Sociální a zdravotní pojištění a SF (FKSP)			364	364
Osobní náklady celkem		1389	1389	1389
Náklady celkem		1455	1493	1575

Náklady z dalších zdrojů předpokládané za celou dobu jeho řešení

Podpora z jiných tuzemských veřejných zdrojů (z jiné kapitoly státního rozpočtu nebo rozpočtů územních samosprávných celků), pokud existuje	0	0	0
Podpora z ostatních veřejných zdrojů (nepatřících do státního rozpočtu nebo rozpočtů územních samosprávných celků), pokud existuje (veřejné zdroje v ČR i v zahraničí)	0	0	0
Podpora z neveřejných zdrojů (zahraniční zdroje, neveřejné tuzemské zdroje, vlastní neveřejné zdroje, pokud existuje	0	0	0



Grantová agentura České republiky

Příloha k B-rozpis

Specifikace a zdůvodnění požadavků pro 1. rok řešení

Uchazeč: **Univerzita Karlova v Praze - Filozofická fakulta**
 Navrhovatel: **doc. RNDr. Vladimír Petkevič, CSc.**

Registrační číslo:
P406/10/0434

Příloha k B-rozpis je nedílnou součástí návrhu projektu a obsahuje v souladu s ustanovením ZD 4.2.7 specifikaci a zdůvodnění každého požadavku uvedeného v **B -rozpis** a **B-osobní náklady**.

Věcné náklady

spotřební materiál (papíry, CD, DVD)	5.000
drobný hmotný majetek (klávesnice, myši)	5.000
SW licence	5.000

Osobní náklady

Vladimír Petkevič	80.000
Milena Hnátková	80.000
Tomáš Jelínek	170.000
Alexandr Rosen	80.000
Hana Skoumalová	80.000
Jiří Hana	400.000
Jiřina Kovaříková	55.000



Grantová agentura České republiky

Část B-osobní náklady

Osobní náklady pro uchazeče pro 1. rok řešení

(finanční údaje se uvádějí jako celočíselné hodnoty v tis. Kč)

Uchazeč: **Univerzita Karlova v Praze - Filozofická fakulta**
 Navrhovatel: **doc. RNDr. Vladimír Petkevič, CSc.**

Registrační číslo:
P406/10/0434

Náklady na mzdy a odměny odborných pracovníků:

Navrhovatel a spolupracovníci:		Pracovní kapacita na řešení projektu 1)	Požadavky na mzdy a odměny od GA ČR
Jméno	Příjmení		
Vladimír	Petkevič	100	80
Tomáš	Jelínek	100	170
Milena	Hnátková	100	80
Alexandr	Rosen	100	80
Hana	Skoumalová	100	80
programátor nebo	počítačový lingvista	100	400

Mzdy a odměny technických a administrativních pracovníků

	Požadavky na mzdy a odměny od GA ČR
Celkový počet technických a administrativních pracovníků:	1 55

Ostatní osobní náklady (na základě dohod o provedení práce nebo dohod o pracovní činnosti):

Typ činnosti (pracovní náplň), popřípadě jméno studenta	Požadavky
ruční anotace	50
konzultace	30

1) V procentech odpovídajících rozsahu úvazku zaměstnanců na řešení grantového projektu.



Grantová agentura České republiky
Czech Science Foundation

Part A

Project proposal

Date of application: 21.04.2009	Panel: P406	Registration No.: P406/10/0434
---	-----------------------	--

Applicant

Surname and name:	Institution (name and address):
doc. RNDr. Vladimír Petkevič, CSc.	Faculty of Arts, Charles University in Prague

Project title:
Syntactic annotation of Czech corpora
Duration (years): 3 Starting date: 01/01/2010

Abstract (max. 15 lines):
The project's aim is to create a set of tools for syntactic annotation of Czech text corpora. Currently, there exist tools for morphological tagging, which were already used for tagging the Czech National Corpus. There are also several systems for syntactic analysis of Czech, but they are heavily dependent on the underlying syntactic theory, and thus their results are not suitable for the purpose of tagging corpora aimed at the general public. Beside it, their success rate is not high enough for fully automatic processing of texts. The proposed project will exploit all of the abovementioned tools. After converting the results to a format that will be syntactically more neutral, we expect a lower error ratio. Furthermore, we will also create a set of tools for error correction and then we will annotate an experimental corpus of the size ca 1 million words.



Grantová agentura České republiky
Czech Science Foundation

Part B

Financial Plan

(In thousands of CZK)

Registration No.: **P406/10/0434**

Applicant: doc. RNDr. Vladimír Petkevič, CSc.	1 st year	2 nd year	3 rd year
Consumables	66	104	186
Equipment	0	0	0
Personnel	1389	1389	1389
Total costs	1455	1493	1575