

# Language Recognizer View

Aplikace slouží k zobrazení výsledků identifikace jazyka zvoleného textu nebo souboru. Na konsoly, případně do souboru, vypíše věty a hodnoty jejich pravděpodobností v jednotlivých modelech. Hodnota s největší pravděpodobností je pak barevně zvýrazněna (identifikovaný model jazyka).

## Povinné parametry

Parametr	Popis
<b>file</b>	Soubor s texty, je možné zadat pouze jeden soubor.
<b>text</b>	Je možné zadat pouze text (v uvozovkách), na kterém se má identifikovat jazyk, pak není povinný parametr „file“
<b>model</b>	Model, který se použije při identifikaci jazyka a je umístěn ve složce „Models“

## Volitelné parametry

Parametr	Popis	Možné hodnoty	Defaultní hodnota
<b>models all</b>	Použije všechny modely ze složky „Models“, která je v hlavní složce aplikace.		
<b>order</b>	Stupeň n-gramů, který se použije pro identifikaci jazyka. Pokud se zvolí vyšší stupeň než mají modely, použije se nejvyšší stupeň modelů.	Celé kladné číslo	5
<b>encoding</b>	Kódování vstupního textového souboru. Stejně kódování se použije pro výstupní soubory.	Textový řetězec	UTF-8
<b>separators</b>	Oddělovače vět v odstavci. Oddělování odstavců je automatické. Pokud se nezadají žádné oddělovače, bude text rozdělen jen podle odstavců.	Textový řetězec	
<b>min_length</b>	Minimální délka vět včetně oddělovače. [znaky]	Celé kladné číslo	1
<b>ignore_next_separator</b>	Pokud je délka věty kratší než parametr „min_lenght“, ignoruje následující oddělovač, tak aby získal delší větu.	true/false	False
<b>lower_case</b>	Převede všechny znaky rozpoznávaného textu na malá písmena. Při použití tohoto parametru by i modely měly být natrénovány na textech s malými písmeny.	true/false	False
<b>file_out</b>	Při nastavení se výsledky nezapisují na konsoly, ale do zadaného souboru.	Textový řetězec	

## Příklad

LanguageRecognizerView file data.txt encoding windows-1250 model Czech model Slovak  
order 6 separators .?! ignore\_next\_separator true

## Výstup

Zobrazení výsledků na konsoly. Vlevo jsou hodnoty  $\log(p)$  zvolených modelů pro rozpoznávanou větu, která je vpravo. Nejvyšší hodnota  $\log(p)$  pro danou větu je pak zvýrazněna barevně.