

Základní informace

Jméno projektu	Analýza sémantiky zdrojového kódu nad platformou Roslyn-PHP kompilátoru pro Visual Studio Code
Zkratka	VsPeachSem
Vedoucí	misek@ksi.mff.cuni.cz
Konzultanti	jakub.misek@iolevel.com , ben.fistein@iolevel.com , zavoral@mff.cuni.cz , husak.robert@seznam.cz
Anotace	Cílem projektu je navrhnout a vytvořit níže specifikovanou analýzu zdrojového kódu jazyka PHP za použití moderní platformy Roslyn a její implementace, která současně vzniká v rámci KSI MFF UK. Součástí práce je integrovat zobrazení výsledků této analýzy do prostředí Microsoft® Visual Studio Code.

Motivace

Jelikož drtivá většina internetu je tvořena dynamickým jazykem PHP, vzniká spousta nástrojů pro údržbu a optimalizaci aplikací na PHP postavených. Několik projektů na MFF UK se již zabývalo jeho různými aspekty, od kompilace [Phalanger, Peachpie] přes prototypování dynamických objektů [Hausar] po bezpečnostní analýzu [Phalanger, Weverca].

Nejnovější přírůstek je kompilátor Peachpie [www.peachpie.io] stavěný nad platformou Microsoft Roslyn. Díky této platformě je možné postupně spojit předchozí snahy pod jedno rozhraní a navázat předchozí výzkum na moderní platformu aplikovatelnou v rozsáhlé škále dalšího použití.

Platforma již poskytuje rozhraní pro implementaci sémantické analýzy a integraci do vývojových prostředí. Pro úspěšné nasazení kompilátoru je potřeba rozšířit a sjednotit analýzu zdrojového kódu (typovou, bezpečnostní a analýzu detekující nekompatibility či nedodělky) a zpřístupnit ji přirozeným způsobem do vývojového prostředí. To položí základ i dalším možným nezávisle definovaným rozšířením sémantické analýzy jinak dynamického jazyka PHP.

Popis projektu

Projekt se skládá ze čtyř funkčních prvků.

- 1) Rozšíření projektu Peachpie o rozhraní pro sémantickou analýzu a reporting jejích výsledků. Projekt staví na dobře zdokumentované platformě Roslyn a již poskytuje datové struktury poskytující sémantický strom a graf řídicích struktur (CFG).
- 2) Sémantická analýza zdrojového kódu pomocí (1) detekující následující:
 - a) Funkcionalitu, která je známa jako zastaralá a nedoporučuje se používat či bude v budoucnu odstraněna. Seznam takovéto funkcionality bude obecně definován v metadatech zdrojových referencí.
 - b) Potencionální běhové chyby založené na typové analýze – type mismatch, unknown method call, unknown function call, missing arguments. Samotná typová analýza je již implementována. V rámci tohoto bodu bude použita.

- c) Konstrukce, které kompilátor aktuálně nepodporuje. Takové konstrukce budou valuovány složitostí pro účely ohodnocení jejich případné implementace. Seznam nepodporovaných konstrukcí musí být aktuální vzhledem k aktuálnímu stavu kompilátoru.
- 3) Vestavba sémantické analýzy (2) do prostředí Visual Studio Code (VSCode). Reportované sémantické chyby budou podtrženy v editoru zdrojového kódu a jejich úplný seznam zobrazen v samostatném okně.
- 4) Podrobná dokumentace rozhraní pro sémantickou analýzu, dokumentace integrace do VSCode a série krátkých článků ilustrující použití platformy v anglickém jazyce.

Platforma, technologie

Backend projektu staví na platformě Microsoft Roslyn, což předurčuje technologie .NET Framework, .NET Core a Mono. Vývojový jazyk je C#.

Frontend pro Visual Studio Code je minimalistický a vyvíjí se v jazycích Javascript či Typescript. Cílovými platformami budou stolní systémy Linux, Windows a Mac OS.

Odhad náročnosti

Počet řešitelů: 4-5

Termín dokončení: 9 měsíců od zahájení

Plán prací: 1. měsíc – detailní analýza požadavků a upřesnění požívaných technologií
 2-3. měsíc – prototyp VSCode integrace, reporting sémantické analýzy
 3.-7. měsíc – implementace analýzy, průběžné psaní dokumentace
 8.-9. měsíc – dokončování a ladění, finální vestavění do VSCode prostředí, finalizace dokumentace

Vymezení projektu

Projekt je zaměřen na následující oblasti (zaškrtněte vyhovující):

Diskrétní modely a algoritmy	
	diskrétní matematika a algoritmy
	geometrie a matematické struktury v informatice
	optimalizace
Teoretická informatika	
	Teoretická informatika
Softwarové a datové inženýrství	
x	softwarové inženýrství
x	vývoj software
	webové inženýrství
	databázové systémy
x	analýza a zpracování rozsáhlých dat

Softwarové systémy	
x	systemové programování
	spolehlivé systémy
	výkonné systémy
Matematická lingvistika	
	počítačová a formální lingvistika
	statistické metody a strojové učení v počítačové lingvistice
Umělá inteligence	
	inteligentní agenti
	strojové učení
	robotika
Počítačová grafika a vývoj počítačových her	
	počítačová grafika
	vývoj počítačových her

Poznámky

žádné