

Zárójelentés szakmai beszámolója

A pályázat címe: „Hatékony rendszer-szintű hatásanalízis módszerek és alkalmazásuk a szoftverfejlesztés folyamatában”

Sorszám: OTKA K-73688

Szeged, 2012. május

A kutatás négyéves időtartama alatt a kutatómunka lényegében terv szerint haladt, kisebb eltérésekkel. Az eredeti tervhez képest új területek is előtérbe kerültek a korábbi eredmények és az eredmények hasznosulása során felmerült egyéb igények alapján. Ezen területekre az alábbi összefoglalóban külön kitérünk.

1. A kutatás első évének főbb eredményei:

A kutatómunka a terület legfrissebb eredményeinek feltérképezésével és egy átfogó tanulmány elkészítésével indult, amelynek eredményeképpen elkészült egy tanulmány, a részterületek rövid ismertetőjével és egy annotált bibliográfiával [BeJ09]. Ezzel párhuzamosan kialakítottunk egy kísérleti környezetet, amelyben a módszerek kipróbálhatóvá, a kísérleti mérések elvégezhetővé váltak. Ennek része egy jelentős mennyiségű forráskódot tartalmazó szoftverbázis, amely forráskód elemzés, dinamikus elemzés és a verziókövető rendszerek adatbányászatára is alkalmas. A módszerek által szolgáltatott eredmények kiértékelésére kialakítottunk egy környezetet és egy támogató szoftvert [FHF08].

A statikus forráskód elemzésen alapuló módszerünkkel (Static Execute After) kapcsolatos kutatásokat intenzíven folytattuk ebben az évben. Megmértük a megközelítés pontosságát valós programokon és ismertettük alkalmazhatóságát nagyméretű szoftverrendszerek elemzésére [JBG08, Jas08]. A kutatási időszakban amerikai kutatótársakkal közösen megvizsgáltuk a forráskód elemek csatolás alapú kiértékelését hatásanalízis céljára. Itt összehasonlítottuk a hagyományos függőség alapú csatolás és információkinyerés alapú megközelítések eredményeit [PMF09]. C/C++ környezetben előfeldolgozóban felmerülő függőségek analízisére kidolgoztunk egy módszert [VJB08, VBG09]. Ebben az évben kezdtük el kutatásainkat a hibára hajlamos szoftverkomponensek detektálásával, hibák előrejelzésével kapcsolatban, hiszen a hatásanalízis egy nagyon fontos alkalmazása megmutatni egy-egy változás kockázatos, hibákat magában rejlő hatásait [Sik08].

2. A második év tevékenységei és eredményei:

Az eredmények hasznosulására nagyobb hangsúlyt fektettünk ebben az évben, olyan kutatásokat is végeztünk, amelyek az alaplómódszerek alkalmazásai irányába is mutatnak, például tesztelés vagy változás-propagáció. Az előző évben elkezdett hibaelőrejelzéssel kapcsolatos kutatásokat is folytattuk.

A korábban tervezett adatbázis technológiákra (SQL) vonatkozó módszerek kutatását ebben az évben végeztük, amelynek eredményeképpen egy nagy, heterogén rendszereket is elemezni képes algoritmus kidolgozásra került [NPG10]. A korábbi eredmények hasznosulásának jelentős területe a szoftvertesztelés. Ennek kapcsán kísérleteket végeztünk tesztesetek generálásával kapcsolatban [BBG09] és megkezdtük kutatásainkat a teszteset szelekció és -priorizálás területén. A hatásanalízis további jelentős területe a változáspropagálás, ezért ezen a területen kísérleteket végeztünk a korábban kidolgozott hatásanalízis módszereink segítségével, hogy meghatározzuk az azok közötti különbségeket. Az algoritmusokat ez esetben nem csak a kísérleti környezetünkben, hanem a programozói tapasztalat figyelembevételével is össze tudtuk hasonlítani. A korábban már elkezdett hiba-előrejelzés területén kidolgoztunk egy nagyméretű alkalmazások dinamikus analízisén alapuló előrejelző módszert [BFG09], valamint folytattuk a különböző

forráskód-jellemzők alapján történő előrejelzéssel kapcsolatos kutatást [HBS09, SiG10]. Ugyancsak születtek eredmények az alacsony szintű konstrukciók (előfeldolgozó, szkriptek) elemzésével kapcsolatban [HFF09, Vid09].

A kutatócsoportban, ebben az évben a jelen kutatási témához kapcsolódóan több doktori disszertáció is készült (Siket István, Gergely Tamás, Jász Judit, Vidács László), többek között a kutatási projektben résztvevő két kutatónké is.

3. A kutatás harmadik évének főbb eredményei:

Hasonlóan az előző időszakhoz, ebben az évben is jobban koncentráltunk az alkalmazásokra, nevezetesen fontos szerepet kapott a tesztelés és általában a szoftverminőségi alkalmazások úgy, mint költségbecslés, hiba előrejelzés, architekturális függőségek elemzése, valamint folytattuk az előző évben megkezdett szakterületi kutatásokat.

A tesztelési módszertanok, tesztlefedettség-mérés és teszt szelekció/priorizálás területén értünk el eredményeket, amik a hatásanalízis fontos alkalmazási területei [GBG10]. A másik kapcsolódó eredmény a különböző szoftver-jellemzők vizsgálata a tesztelhetőség szempontjából, valamint a minőségjavító változtatások (refactoring) és azok hatásanalízise [HHH10]. A hiba-előrejelzés területén végzett kutatásokat ebben az évben is folytattuk [Fer10]. A változáspropagálás területén végzett kísérleteinket kiterjesztettük élő programozói tapasztalat figyelembevételével, hiszen ezen hatásanalízis módszereknek tipikusan a fejlesztők a fő használói [THJ10]. Az előrejelzés további fajtája a költségek becslése, ami a változtatás hatásainak vizsgálata mellett a változtatás várható ráfordításának vizsgálatát is igényli. Ezen a területen elindítottunk kutatásainkat, melynek célja komplex előrejelző modellek megalkotása, amely figyelembe veszi a kód és a folyamat jellemzőit is. A rendszer szintű függőségek kezelése területén a komponensek közötti függőségek megfelelő meghatározásával és ábrázolásával foglalkoztunk, ezek vizualizálási kérdéseit vizsgáltuk, hiszen nagyméretű, komplex rendszerek esetében ez nagy segítség lehet [SFF10]. Kapcsolódó kutatás az ún. fogalmi csatolás és fogalmi metrikák vizsgálata, ami ugyancsak a rendszer szintű, osztott rendszerek esetén alkalmazható sikerrel [UFP10]. Kutatásaink hatására már ebben az időszakban nagyméretű adatbázis központú rendszerekre kidolgozott módszereket kezdtek el alkalmazni [NVF10].

Mivel a tesztelést az egyik legfontosabb alkalmazási területnek tekintettük továbbra is, ebben az időszakban tudatosan elkezdtük ismereteinket bővíteni ezen a területen. Két munkatársunk, Beszédes Árpád és Gergely Tamás megszerezték a haladó szintű szoftvertesztelői képesítést (ISTQB Certified Tester Advanced Level Test Analyst). A tapasztalatokat elkezdtük beépíteni a képzési rendszerünkbe is, amely eredményeképpen új tárgyak keretén belül vezettük/vezetjük be a szoftvertesztelés oktatását.

4. A kutatás negyedik, záró évének eredményei:

Hasonlóan a korábbi évekhez, ebben az időszakban is főleg a módszerek gyakorlati alkalmazásai irányában végeztük kutatásainkat, valamint igyekeztük az elért eredményeinket összefoglaló jellegű tanulmányokban közzé tenni.

Nagyméretű adatbázis központú rendszerek kezelésére is alkalmas elemzési módszereink véglegesedtek (minőség elemzés és teszt-automatizálás), és ennek hatására azokat elkezdték alkalmazni speciális Magic környezetben [FNF11, NVF11b]. A nagyméretű rendszerek elemei közötti kapcsolatok és egyéb metrikákkal kapcsolatos kutatásainkat is folytattuk, mint a hatásanalízis egy fontos alkalmazási területét. A metrikák alkalmazásával létrehoztunk egy minőségi jellemzőket leíró modellt [BHK11, BHL12, SFB12,

HBI11]. Produktivitás előrejelzés területén is értünk el új eredményeket, megmutattuk, hogy nem csak a forráskódból kiolvasható összefüggések, de a projekt folyamattal kapcsolatos attribútumok is fontos szerepet játszanak a sikeres előrejelzésnél [TVB11a, TVB11b].

Erre az időszakra az alapvető hatásanalízis eszközeink már olyan szintre fejlődtek, hogy igazán nagyméretű, valós alkalmazások elemzésére gond nélkül alkalmazhatók. A legújabb terület, amellyel a kutatás utolsó évében kezdtünk el foglalkozni a WebKit nevű, nyílt forráskódú alkalmazás. Ez a rendszer több millió programsorból és több tízezer tesztelésből áll, valamint nagyon intenzív fejlesztés szereplője. Mindezek alkalmassá teszik arra, hogy a kísérleteink központi szereplőjévé váljon, és ezért a jelen kutatás lezárulta után is ezen rendszerrel szeretnénk folytatni a munkánkat.

Ebben a környezetben megvalósítottunk egy regressziós teszt szelekciós módszert, felhasználva a korábbi eredményeinket, és a módszerhez tartozó prototípust a WebKit felhasználói közösség éles üzemben elkezdte használni. Az eredményeinkről írt publikáció jelenleg bírálat alatt áll [BGS12]. Hatásanalízis módszereinket, főleg a Static Execute After algoritmust is megvalósítottuk ebben a környezetben, és megvizsgáltuk az algoritmus sikerességét a tényleges hibák megtalálásának tekintetében. Pontosabban, azt vizsgáltuk, hogy a WebKit fejlesztése során ténylegesen tapasztalt regressziós hibák esetében az azokat okozó változtatásokra számított hatáshalmazok milyen arányban találják el a tényleges javításhoz szükséges módosításokat. Eredményeink azt mutatták, hogy az esetek többségében sikeres a hatásanalízis. Az eredmények első részét már ismertettük [JSB12], míg a részletesebb eredményeket leíró tanulmányunk jelenleg bírálat alatt áll [SJG12].

Ebben az időszakban zártuk le a dinamikus szelekteléssel és annak hatásanalízis alkalmazásaival kapcsolatos korábbi kutatásainkat is. Kidolgoztunk egy keretrendszert az összes kapcsolódó algoritmus közös kezelésére, amely eredményeket egy folyóiratban szeretnénk publikálni. Jelen pillanatban a kézirat bírálat alatt áll [BGG12].

5. **Konklúzió:**

A kutatás eredeti célja olyan hatásanalízis módszerek kidolgozása volt, melyek hatékonyan és sikeresen alkalmazhatók nagyméretű, valós alkalmazások esetében is. A célt a kutatás során teljesítettük, mivel a kidolgozott módszerek alapját képezték további projekteknek, melyek során szoftvertermékeket is kiegészítettek módszereink alapján. A kutatás kezdetén zömében a meglévő algoritmusok (statikus és dinamikus programszelektelés és függőség elemzés) továbbfejlesztésén dolgoztunk, de hamarosan világossá vált, hogy bizonyos kapcsolódó területekre (függőségek metrikákkal történő vizsgálata, fogalmi csatolás kutatása, minőségi modellek, hiba- és produktivitás előrejelzés) is ki kell terjesztenünk a kutatást, amelyek a valós alkalmazásokat leginkább támogatják. Speciális technológiákra koncentrálna új szerű eredmények születtek, például adatbázis rendszerek vagy alacsony szintű nyelvek esetében. A kutatás második felében figyelmünket a hatásanalízis módszerek alkalmazásaira fektettük, különösen a tesztelés optimalizálása, teszt lefedettség mérés, -priorizálás és változás propagálás területeken. Eredményeinket nagyrészt a szakterület legjelentősebb nemzetközi konferenciáin ismertettük (konferencia legjobb cikke díját két esetben mi nyertük el), ahol alkalmunk adódott a kapcsolódó külföldi kollégákkal való gondolatcserére is (ennek következtében közös publikációk is születtek), de folyóirat publikációkat is készítettünk. A kutatás eredményeképpen a kutatásban résztvevők és a tágabb kutatócsoport jelentős tudásra tett szert, ami megalapozza az ez irányú kutatások folytatását. Jelentős mértékben növekedett a nemzetközi kapcsolataink száma, számos konferencián vettünk részt, továbbá magunk is szerveztünk konferenciát, illetve kutatóink különböző programbizottsági tagságokkal rendelkeztek. Két résztvevő kutatónk időközben PhD fokozatot szerzett a szakterülethez kapcsolódó eredményeik alapján.

6. Hivatkozások:

1. kutatási év

[JBG08]

Judit Jász, Árpád Beszédes, Tibor Gyimóthy, Václav Rajlich: **Static Execute After/Before as a Replacement of Traditional Software Dependencies**, pp. 137-146 in Proceedings of the 2008 IEEE International Conference on Software Maintenance (ICSM'08), IEEE Computer Society, 2008.

[VJB08]

László Vidács, Judit Jász, Árpád Beszédes, Tibor Gyimóthy: **Combining Preprocessor Slicing with C/C++ Language Slicing**, pp. 163-171 in Proceedings of the 16th IEEE International Conference on Program Comprehension (ICPC'08), IEEE Computer Society, 2008. (*A konferencia legjobb cikke díj.*)

[FHF08]

Fülöp Lajos Jenő, Hegedűs Péter, Ferenc Rudolf: **BEFRIEND – a Benchmark for Evaluating Reverse Engineering Tools**, Periodica Polytechnica (52/3-4), pages 153-162, Budapesti Műszaki és Gazdaságtudományi Egyetem, 2008.

[Jas08]

Jász Judit: **Static Execute After Algorithms as Alternatives for Impact Analysis**, Periodica Polytechnica (52/3-4), pages 163-176, Budapesti Műszaki és Gazdaságtudományi Egyetem, 2008.

[Sik08]

Siket István: **Evaluating the Effectiveness of Object Oriented Metrics for Bug Prediction**, Periodica Polytechnica (52/3-4), pages 177-186, Budapesti Műszaki és Gazdaságtudományi Egyetem, 2008.

[BeJ09]

Beszédes Árpád, Jász Judit: **Tanulmány hatásanalízis módszerekről**, Szegedi Tudományegyetem, 2009.

[PMF09]

Denys Poshyvanyk, Andrian Marcus, Rudolf Ferenc, Tibor Gyimóthy: **Using information retrieval based coupling measures for impact analysis**, Empirical Software Engineering, Springer Netherlands, Volume 14, Number 1: 5-32, 2009

[VBG09]

László Vidács, Árpád Beszédes, Tibor Gyimóthy: **Combining Preprocessor Slicing with C/C++ Language Slicing**, Science of Computer Programming, volume 74, number 7: 399 - 413, 2009

2. kutatási év

[BFG09]

Árpád Beszédes, Lajos Jenő Fülöp and Tibor Gyimóthy: **Predicting Critical Problems from Execution Logs of a Large-Scale Software System**, in Proceedings of the 11th Symposium on Programming Languages and Software Tools and 7th Nordic Workshop on Model Driven Software Engineering, pages 19-30 (SPLST 2009, Finland, Tampere, August 26-28, 2009).

[BBG09]

Tibor Bakota, Árpád Beszédes, Tamás Gergely, Milán Gyalai, Tibor Gyimóthy and Dániel Füleki: **Semi-Automatic Test Case Generation from Business Process Models**, in Proceedings of the 11th Symposium on Programming Languages and Software Tools and 7th Nordic Workshop on Model Driven Software Engineering (SPLST 2009 & NW-MODE 2009), pages 5-18. Tampere, Finland, August 26-28, 2009.

[HFF09]

György Hegedűs, Lajos Jenő Fülöp and Rudolf Ferenc: ***Script2Ant - a tool for cross-platform scripting between different operating systems***, in Proceedings of 11th Symposium on Programming Languages and Software Tools and 7th Nordic Workshop on Model Driven Software Engineering, pages 45-58 (SPLST 2009, Finland, Tampere, August 26-28, 2009).

[HBS09]

Maarit Harsu, Tibor Bakota, István Siket, Kai Koskimies, and Tarja Systä: ***Code Clones: Good, Bad, or Ugly?***, in Proceedings of the Proceedings of 11th Symposium on Programming Languages and Software Tools and 7th Nordic Workshop on Model Driven Software Engineering, pages 31-43 (SPLST 2009, Finland, Tampere, August 26-28, 2009).

[Vid09]

László Vidács: ***Refactoring of C/C++ Preprocessor Constructs at the Model Level***, in Proceedings of the 4th International Conference on Software and Data Technologies (ICSOFT'09), pages 232-237. Sofia, Bulgaria, July 26-29, 2009, INSTICC Press.

[SiG10]

István Siket and Tibor Gyimóthy: ***The Software Developers' View on Product Metrics; A Survey-based Experiment***, Annales Mathematicae et Informaticae, Vol. 37, pages 225-240, Líceum University Press, Eger, 2010.

[NPG10]

Csaba Nagy, János Pántos, Tamás Gergely and Árpád Beszédes: ***Towards a Safe Method for Computing Dependencies in Database-Intensive Systems***, in Proceedings of the 14th European Conference on Software Maintenance and Reengineering (CSMR 2010), pages 169-178. Madrid, Spain, March 15-18, 2010, IEEE Computer Society.

3. kutatási év

[THJ10]

Gabriella Tóth, Péter Hegedűs, Judit Jász, Árpád Beszédes and Tibor Gyimóthy: ***Comparison of Different Impact Analysis Methods and Programmer's Opinion - an Empirical Study***, in Proceedings of the 8th International Conference on the Principles and Practice of Programming in Java (PPPJ 2010), pages 109-118. Vienna, Austria, September 15-17, 2010.

[GBG10]

Tamás Gergely, Árpád Beszédes, Tibor Gyimóthy and Milán Gyalai: ***Effect of Test Completeness and Redundancy Measurement on Post Release Failures - an Industrial Experience Report***, in Proceedings of the 2010 IEEE International Conference on Software Maintenance (ICSM 2010), pages 1-10. Timisoara, Romania, September 12-18, 2010, IEEE Computer Society.

[Fer10]

Rudolf Ferenc: ***Bug Forecast: A Method for Automatic Bug Prediction***, in Proceedings of the 2010 International Conference on Advanced Software Engineering & Its Applications (ASEA 2010), International Convention Center Jeju, Jeju Island, Korea, pages 284-296, December 13-15, 2010. Published in Lecture Notes in Computer Science, Springer-Verlag.

[HHH10]

György Hegedűs, György Hrabovszki, Dániel Hegedűs and István Siket: ***Effect of Object Oriented Refactorings on Testability, Error Proneness and other Maintainability Attributes***, in Proceedings of the 1st Workshop on Testing Object-Oriented Systems (ETOOS '10). Maribor, Slovenia, June 21-25, 2010.

[SFF10]

Lajos Schrettner, Lajos Jenő Fülöp, Rudolf Ferenc and Tibor Gyimóthy: **Visualization of Software Architecture Graphs of Java Systems**, in Proceedings of the 8th ACM International Conference on the Principles and Practice of Programming in Java (PPPJ 2010), Vienna, Austria, pages 148-157, September 15-17, 2010.

[UFP10]

Béla Újházi, Rudolf Ferenc, Denys Poshyvanyk and Tibor Gyimóthy: **New Conceptual Coupling and Cohesion Metrics for Object-Oriented Systems**, in Proceedings of the 10th IEEE International Working Conference on Source Code Analysis and Manipulation (SCAM 2010), Timisoara, Romania, pages 33-42, September 12-13, 2010. (*A konferencia legjobb cikke díj.*)

[NVF10]

Csaba Nagy, László Vidács, Rudolf Ferenc, Tibor Gyimóthy, Ferenc Kocsis and István Kovács: **MAGISTER: Quality Assurance of Magic Applications for Software Developers and End Users**, in Proceedings of the 26th IEEE International Conference on Software Maintenance (ICSM'10), Tool Demos Track, pages 1-6, Timisoara, Romania, September 12-18, 2010.

4. kutatási év

[FNF11]

Daniel Fritsi, Csaba Nagy, Rudolf Ferenc and Tibor Gyimóthy: **A Layout Independent GUI Test Automation Tool for Applications Developed in Magic/uniPaaS**, in Proceedings of the 12th Symposium on Programming Languages and Software Tools (SPLST'11), pages 248-259, Tallinn, Estonia, October 5-7, 2011.

[NVF11b]

Csaba Nagy, László Vidács, Rudolf Ferenc, Tibor Gyimóthy, Ferenc Kocsis and István Kovács: **Complexity Measures in 4GL Environment**, in Proceedings of the 2011 Computational Science and Its Applications (ICCSA'11), Santander, Spain, June 20-23, 2011. Published in Lecture Notes in Computer Science (LNCS), Springer-Verlag, volume 6786, pages 293-309, 2011.

[BHK11]

Tibor Bakota, Péter Hegedűs, Péter Körtvélyesi, Rudolf Ferenc and Tibor Gyimóthy: **A Probabilistic Software Quality Model**, in Proceedings of the 27th IEEE International Conference on Software Maintenance (ICSM'11), Williamsburg, VA, USA, pages 243-252, September 25-30, 2011.

[HBI11]

Péter Hegedűs, Tibor Bakota, László Illés, Gergely Ladányi, Rudolf Ferenc and Tibor Gyimóthy: **Source Code Metrics and Maintainability: a Case Study**, in Proceedings of the 2011 International Conference on Advanced Software Engineering & Its Applications (ASEA'11), Jeju Grand Hotel, Jeju Island, Korea, December 8-10, 2011, pages 272-284. Published by Springer-Verlag.

[TVB11a]

Gabriella Tóth, Ádám Zoltán Végh, Árpád Beszédes and Tibor Gyimóthy: **Adding Process Metrics to Enhance Modification Complexity Prediction**, in Proceedings of the 19th IEEE International Conference on Program Comprehension (ICPC'11), pages 201-204. Kingston, Ontario, Canada, June 22-24, 2011, IEEE Computer Society.

[TVB11b]

Gabriella Tóth, Ádám Zoltán Végh, Árpád Beszédes, Lajos Schrettner, Tamás Gergely, and Tibor Gyimóthy: **Adjusting Effort Estimation Using Micro-Productivity Profiles**, in Proceedings of the 12th

Symposium on Programming Languages and Software Tools (SPLST'11), pages 207-218. Tallin, Estonia, October 5-7, 2011.

[BHL12]

Tibor Bakota, Péter Hegedűs, Gergely Ladányi, Péter Körtvélyesi, Rudolf Ferenc, and Tibor Gyimóthy: *A Cost Model Based on Software Maintainability*, bírálóat alatt, benyújtva a 28th IEEE International Conference on Software Maintenance (ICSM'12) konferenciára.

[SFB12]

Lajos Schrettner, Lajos Jenő Fülöp, Árpád Beszédes, Ákos Kiss and Tibor Gyimóthy: *Software Quality Model and Framework with Applications in Industrial Context*, in Proceedings of the 16th European Conference on Software Maintenance and Reengineering (CSMR'12) , Industrial Track, 2012, pp. 453-456.

[BGS12]

Árpád Beszédes, Tamás Gergely, Lajos Schrettner, Judit Jász, László Langó and Tibor Gyimóthy: *Code Coverage-Based Regression Test Selection and Prioritization in the WebKit System*, bírálóat alatt, benyújtva a 28th IEEE International Conference on Software Maintenance (ICSM'12) konferenciára.

[JSB12]

Judit Jász, Lajos Schrettner, Árpád Beszédes, Csaba Osztrogonác and Tibor Gyimóthy: *Impact Analysis Using Static Execute After in WebKit*, in Proceedings of the 16th European Conference on Software Maintenance and Reengineering (CSMR'12), 2012, pp. 95-104.

[SJG12]

Lajos Schrettner, Judit Jász, Tamás Gergely, Árpád Beszédes and Tibor Gyimóthy: *Impact Analysis in the Presence of Dependence Clusters Using Static Execute After in WebKit*, bírálóat alatt, benyújtva a 12th IEEE International Working Conference on Source Code Analysis and Manipulation (SCAM'12) konferenciára.

[BGG12]

Árpád Beszédes, Tamás Gergely and Tibor Gyimóthy: *Investigation of Different Graph-Less Dynamic Program Slicing Algorithms*, bírálóat alatt, benyújtva a Science of Computer Programming folyóirathoz (Elsevier).