

Munkafolyam-alkalmazások szerkesztésének támogatása csoportmunka-módszerekkel

SIPOS GERGELY, KACSUK PÉTER

MTA SZTAKI

{sipos, kacsuk}@sztaki.hu

Kulcsszavak: zárolás, konzisztencia, workflow, szerkesztés, P-GRADE

A munkafolyam-leírások üzleti-ipari alkalmazások, valamint tudományos számítások modellezésére és számítógéppel segített automatizálására létrehozott megoldások. A cikk ismerteti az irányított gráfokként felfogható munkafolyam-alkalmazások több személy általi szerkesztésének lehetőségét, olyan módszereket, melyekkel folyamatfejlesztői csoportok képesek gyorsan és kis erőfeszítéssel új szimulációkat létrehozni és végrehajtani. Az ismertetésre kerülő megoldás zárolással biztosítja a hozzáférést több ember számára ugyanazon munkafolyam-leírás különböző részeihez, úgy, hogy eközben a gráf tartalmi és formai helyességét is garantált. A helyesség megőrzése kritikus fontosságú ahhoz, hogy a gráf alapján a folyamat valóban végrehajtható legyen. A munka ezen felül röviden ismerteti a módszer P-GRADE Portál munkafolyam környezetben való használatát is.

1. Bevezetés

A mai gyorsan fejlődő, gyorsan változó üzleti és ipari világban egy vállalat sikerének legfontosabb záloga, ha minél alacsonyabb költséggel képes rutinszerű feladatait ellátni, illetve ha a lehető leggyorsabban és leghatékonyabban tud a piac új kihívásaira reagálni. Ennek érdekében a vállalatoknak folyamatosan felülbírálni, javítani, optimalizálni kell a saját belső, illetve őket a környezetbe ágyazó külső folyamataikat. Az 1980-as évek vége, 1990-es évek eleje körül kezdték számítógéppel segíteni az ipari-üzleti folyamatok leírását és automatizálását. A számítógéppel segített folyamatok ekkor kapták a munkafolyam (angolul workflow) elnevezést [4,9].

Egy munkafolyam-alkalmazás olyan irányított gráfokként fogható fel, melyben a csomópontok elvégzendő feladatokat reprezentálnak, az élek pedig a feladatok közötti függőségeket jelentenek. Például egy banki hitellelbírálás folyamatát leíró munkafolyam-alkalmazásban a csomópontok az elbírálás lépéseit, az élek pedig a banki alkalmazottak közötti dokumentumok továbbítását jelenthetik. Munkafolyam-leírások többnyire grafikus szerkesztőrendszerrel készülnek. A munkafolyam létrehozása után, jellegétől függően különböző számítógépes és emberi erőforrásokat és szolgáltatásokat használva kerül végrehajtásra. Egy munkafolyam-alkalmazásokat kezelő rendszer ennek valamilyen szerkesztőkörnyezetéből, a munkafolyamot végrehajtani képes vezérlőegységből, továbbá olyan felületekből és komponensekből áll, amelyekkel a folyamat lépéseit elvégző felhasználók és számítógépes programok a rendszerhez kapcsolódhatnak (az előző példában a banki alkalmazottak illetve a banki adatbázisok.)

Az elmúlt 20 év alatt a munkafolyam-alkalmazások az üzleti-ipari világ mellett számos más körben is elterjedtek és emiatt maga a munkafolyam fogalma erősen

túlterhelte vált. Például számos tudományos terület hívja munkafolyamnak a kutatási, vagy kutatást kiértékelő folyamatokat. Különösen jellemző ez a tudomány számítógépes szimulációs megoldásokra erőteljesen építő ágaiban, az úgynevezett e-tudományokban (elektronikus tudomány) [8].

Bármilyen alkalmazási területet is tekintünk, ugyanúgy igaz, hogy a munkafolyam-alkalmazások egyre bonyolultabb felépítésűek, egyre több tudást tárolnak, egyre összetettebb gráfokkal írhatók le. Ez egyrészt az alkalmazások méretén, másrészt a minőségén is látszik [1]. Az ilyen alkalmazások létrehozása hagyományos, egyetlen felhasználót támogató szerkesztőrendszerekkel időigényes és sok hibalehetőséget ad. Munkafolyam-gráfok hatékony fejlesztéshez csoportmunka-támogatásra van szükség.

A számítógépes csoportmunka, vagy más néven kollaboratív munka támogatása az 1980-as évek közepén az asztali számítógépek és számítógép-hálózatok elterjedésével egyidőben kezdődött. Az első kollaboratív alkalmazások elektronikus naptárak voltak, melyekkel könnyebben lehetett közös találkozókhöz mindenkinek megfelelő időpontot választani [5]. A csoportmunka-kutatás egy erősen interdiszciplináris terület, melyben az informatikusok mellett pszichológusok és szociológusok is tevékenykednek. Míg utóbbiak főleg a csoportok motivációival, módszereinek és elégedettségének analízisével foglalkoznak, az informatikusok csoportmunkát támogatni képes eszközöket terveznek, implementálnak és üzemeltetnek.

A leghatékonyabb csoportmunkát lehetővé tevő eszközök az úgynevezett „valós idejű kollaboratív rendszerek”. Segítségükkel egyetlen számítógépes fájl tud több felhasználó egyidőben szerkeszteni úgy, hogy a rendszer biztosítja az adatok helyességét. Ezen garancia miatt nem lesz elveszett, felülírt munka, továbbá nem keletkezik értelmetlen, értelmezhetetlen fájl tartalom.

A cikk egy olyan módszert ismertet, melynek segítségével munkafolyam-szerkesztő eszközök többfelhasználós, valós idejű kollaboratív fejlesztőkörnyezetté alakíthatók. A módszert az MTA SZTAKI Párhuzamos és Elosztott Rendszerek Kutatólaboratóriumának munkatársai dolgozták ki, akik ez alapján egy saját fejlesztésű munkafolyam-szerkesztő és -futtató eszköznek, a P-GRADE Grid Portálnak létrehozták a csoportmunkát elősegítő változatát. A környezettel nemzetközi tudományos csoportoknak segítenek számításigényes és nagy adathalmazokat feldolgozó munkafolyam-alkalmazások létrehozásában.

2. Gráf fejlesztés támogatása csoportmunka-módszerekkel

Valós idejű csoportmunka-támogató rendszerekben az interaktív (alacsony válaszdíjt biztosító) felhasználói élmény megtartása az egyik legfontosabb követelmény. A fejlesztőkörnyezetnek a helyi felhasználó akcióira gyorsan kell reagálnia (mintha továbbra is egyfelhasználós rendszerrel lenne dolga), ugyanakkor a távoli felhasználók által végzett módosításokat is a lehető leghamarabb kell megjelenítenie. Ez a követelmény a csoportmunka-rendszerekben egy olyan, úgynevezett „replikált architektúrát” eredményez, melyben a megosztott fájl – esetünkben a munkafolyam gráfja – a fejlesztőcsoport minden tagjának gépén egy-egy példányban tárolódik.

A replikált architektúra lehetőséget ad a felhasználóknak a saját gépükön tárolt gráf gyors módosítására, ezen változtatások saját felületen való azonnali megjelenítésére. Mivel azonban a fejlesztők interneten keresztül összekapcsolt gépei között az adatátvitel ideje nem nulla, ezért a gráfmódosítási üzenetek a különböző gépekre más-más sorrendben érkehetnek meg, és különböző sorrendben kerülhetnek végrehajtásra. Ez az egyes gépeken a munkafolyam gráfjának különböző és helytelen változatainak a létrejöttét eredményezheti.

Csoportmunka-támogató rendszerekben a megosztott fájl konzisztenciáját sértő helyzetek ellen konkurenciakezelő megoldásokkal védekeznek [2]. A legelterjedtebb konkurenciakezelő megoldások a gépek között átküldött üzenetek mesterséges késleltetésén, vagy az üzenetek átalakításán alapulnak. Mesterséges késleltetéssel megoldható, hogy minden gépre azonos sorrendben jussanak el a módosítási utasítások, így minden gépen ugyanaz a fájl tartalom jön létre. Alternatív módon az üzenetek átalakításával nem megfelelő sorrendű beérkezés esetén is biztosítható, hogy ugyanaz a fájl tartalom keletkezzen minden gépen.

Az általunk gráfok szerkesztésére javasolt módszer zárolás segítségével képes felosztani a gráfot, majd annak részeit különböző felhasználók számára allokálni [7]. A zárolás biztosítja, hogy a munkafolyam-gráfon dolgozó személyek közül egyidőben többen ne tudják ugyanazt a csomópontot, vagy ugyanazt az élt módosítani. Ez a módszer lehetővé teszi a csoportos szerkesztést –

mivel a gráf különböző részei más-más felhasználók számára egyidőben válnak elérhetővé – továbbá biztosítja, hogy a fejlesztők lássák és értesülhessenek egymás módosításairól.

A gráf csomópontjainak a zárolása egy központi szolgáltatás feladata, amely kapcsolatban áll az összes résztvevő számítógépével. A felhasználók a saját gépükön futó szerkesztőrendszer felületén keresztül nyomon követhetik a csomópontok és élek foglaltságát, lefoglalásra részgráfokat igényelhetnek a még szabad részekből, továbbá módosíthatják, kiterjeszthetik vagy éppen törölhetik a saját komponenseiket. A rendszer gondoskodik a gépek közötti kapcsolatok fenntartásáról, a különböző gépeken kezdeményezett változások többi gép számára való továbbításáról és a megfelelő hozzájárulásszabályozásról. Mivel a zárolással egymást kiegészítő részekre oszthatók a munkafolyam-gráfok, ezért a gépek közötti kommunikációs késleltetés nem okoz problémát. Ezzel a módszerrel sem mesterséges késleltetések beiktatására, sem az utasítások átalakítására nincs szükség.

3. Munkafolyam-gráfok helyességének megőrzése

Valós idejű kollaboratív szerkesztőrendszerek megalkotása közben felmerülő egyik legkomolyabb probléma a megosztott adatfájl helyességének a biztosítása. A minden gépen azonos fájl tartalom önmagában még nem biztosítja ezt, ugyanis ez csak formai, de nem tartalmi helyességet eredményez. Tartalmi helyesség hiányában előfordulhat, hogy minden gépen ugyanaz a munkafolyam-leírás jött létre, a leírás minden gépen tartalmazza az összes fejlesztő által végzett változtatást, viszont ez a változat egyszerűen nem értelmezhető, nem hajtható végre. Ez annak lehet a következménye, hogy a felhasználók egyidőben történt módosításai jelentésszerű anomáliákat adnak a gráfhoz.

Gráfok esetén az alábbi három ilyen anomáliát azonosítottunk:

1. Egy felhasználó olyan csomópontot töröl, melyhez élek kapcsolódnak, és ezen éleken épp mások módosításokat végeznek. A törlés miatt az él sehova sem vezet, felesleges állé válik, azonban annak törlése az élek szerkesztését végző felhasználók számára elvégzett munkájuk elvesztését jelentené.

2. Egyszerre több felhasználó is bejövő élt akar kapcsolni ugyanazon csomópont bejövő adatokat fogadó csatornájához. Az ugyanoda mutató élek a munkafolyam végrehajtása közben adatvesztést eredményeznének.

3. A munkafolyamhoz egyidőben több felhasználó is új éleket ad, és ezen élek az irányított gráfban körutat eredményeznek. Körutak feloldására a munkafolyam-futtató rendszerek többsége nem képes.

Ennek a három tartalmi anomáliának a megelőzésére a kutatócsoportunk speciális gráf zárolási mechanizmusokat hozott létre. A módszerek képesek a fel-

használók zárolási kérései valamint a munkafolyam-gráf aktuális szerkezete alapján olyan lefoglalható rész-gráfokat meghatározni, melyek zárolása és módosítása nem rontja el a gráf tartalmi helyességét. A felhasználók tetszőleges módosításokat végezhetnek az így lefoglalt részgráfjaikon belül, bármilyen komponens módosíthatnak, létrehozhatnak és törölhetnek, akkor is biztosak lehetnek abban hogy az új változat helyes és futtatható lesz.

A csoportunk öt különböző, a gráfok tartalmi helyességét is figyelembe venni képes zároló algoritmust hozott létre. Az algoritmusok különböző gráfstruktúrákon képesek a zárolási részek meghatározására, és emiatt más-más felépítésű munkafolyam alkalmazásokra alkalmazhatók.

4. A módszer alkalmazása

Munkafolyam-programozó és -futtató technológiák a felhasználók egyre szélesebb rétege számára válnak elérhetővé. Kollaboratív módon, több felhasználó által létrehozott munkafolyamok hasznosságáról ugyan több cikk is szól – többek között az anyagtudomány, orvosi diagnosztizálás, krízismenedzsment vagy mérnöki szimulációk területéről – ugyanakkor az általunk végzett, és

a cikkben bemutatott munka az első részletes analízis, rendszerterv illetve megvalósítás egy ilyen környezetre.

A leírt megoldás az MTA SZTAKI által készített „P-GRADE Portál” rendszerben került megvalósításra (lásd az ábrát) [6]. A P-GRADE Portál lehetővé teszi nagy számítási kapacitást igénylő szimulációk munkafolyam-gráfjainak létrehozását, és több ezer vagy akár több tízezer számítógépből álló, úgynevezett grid-rendszeren való végrehajtását. A P-GRADE Portál által kezelt munkafolyamokban a csomópontok egy-egy számítási feladatok reprezentálnak, a közöttük futó élek pedig a számítások közötti adatfüggőségeket jelentik. (Amennyiben egy A és egy B csomópont között irányított él fut, akkor az A algoritmus által létrehozott adatra a B algoritmusnak a futásakor szüksége van.) A P-GRADE Portál a munkafolyamok futtatása közben biztosítja az algoritmusok közötti adatátvitelt, továbbá garantálja az algoritmusok lefuttatását nagyszámú számítógépet tartalmazó grid-rendszereken.

Kutatócsoportunk 2007 óta az EU által támogatott, nemzetközi „Enabling Grids for E-Science” projekt keretén belül a P-GRADE Portál segítségével több mint 20 szimulációs alkalmazást hozott létre hazai és külföldi alkalmazói csoportok számára [3]. Az alkalmazók köre a kémikusoktól kezdve, a logisztikai szakembereken át a biológusokig széles kört ölel fel.

Munkafolyam-gráf a P-GRADE Portál szerkesztő (jobbra) és végrehajtó rendszerében (balra)

The screenshot displays the P-Grade Portal interface. On the left, the 'Workflow Manager' shows a table of workflows with their status and size. On the right, the 'Workflow Editor' shows a diagram of a workflow graph with nodes and connections.

Workflow	Status	Size
080225_fem25d_ps_gen_coll_ha	incomplete	1.911 [MB]
080225_fem25d_ps_gen_coll_ha_farkasz	init	616.352 [KB]
080225_fem25d_ps_gen_coll_ha_totha	init	616.731 [KB]
0_Rod_Traffic-simulation_workflow	incomplete	956.896 [KB]
1_det2x2	init	204.702 [KB]
2_Matrix_27_hatvany	init	122.137 [KB]

The workflow editor shows a graph with nodes labeled 'INIT', 'SEQ', 'LM_P', 'LM_S_1' through 'LM_S_6', and 'LM_P_2'. The nodes are connected by arrows, representing the flow of data and computation between different stages of the workflow.

5. Összefoglalás

A cikk irányított gráfként modellezhető munkafolyam-alkalmazások csoportmunka-módszerekkel való szerkesztéséhez ismertetett egy módszert. A zár alapú hozzáférés-szabályozás hatékony, mégis ellenőrzött hozzáférést biztosít a fejlesztésben részt vevő csoport tagjai számára. A csoportmunka-támogatással kiegészített P-GRADE Portál a szimulációs alkalmazások fejlesztéséhez nyújt megoldást. A rendszer megkönnyíti a fejlesztőknek egymással, valamint a budapesti székhelyű intézményünkkel való kapcsolattartását.

A szerzőkről



SIPOS GERGELY 2003 óta dolgozik az MTA SZTAKI Párhuzamos és Elosztott Rendszerek Laboratóriumában. MSc fokozatát 2003-ban a Miskolci Egyetemen szerezte, ahol jelenleg doktori tanulmányait is végzi. Kutatási témája munkafolyam-alapú alkalmazások csoportmunka-módszerekkel való kiterjesztése. Szakértője a szolgáltatásalapú grid-rendszereknek, a magas szintű grides felhasználói környezeteknek és a csoportmunka-eszközöknek. Közel 50 konferencia-cikke, több folyóiratcikke és könyvfejezete jelent meg ezekben a témákban. 2003 óta számos magyarországi és EU kutatás-fejlesztési projektben vett részt. 2005 óta az EGEE Enabling Grids for E-sciencE (EGEE), EGEE-II, EGEE-III, valamint European Grid Initiative (EGI) projektekben végez oktatói és alkalmazás támogatói feladatokat.



KACSUK PÉTER az MTA SZTAKI tudományos tanácsadója és a Párhuzamos és Elosztott Rendszerek Laboratórium vezetője. MSc fokozatát és egyetemi doktori címét a Budapesti Műszaki Egyetemen kapta 1976-ban, illetve 1984-ben. Az MTA-tól a Műszaki Tudományok kandidátusa címet 1989-ben kapta. A Bécsi Egyetemen habilitált 1997-ben. A professzori címet 1999-ben kapta meg, az Akadémia Doktora (DSc) címet pedig 2001-ben. Mellékállású professzor az ELTE-n és a Westminsteri Egyetemen. Vendégkutatóként, vagy vendégprofesszorként dolgozott Ausztriában, Angliában, Németországban, Spanyolországban, Ausztráliában és Japánban. Két könyvet, két egyetemi jegyzetet és több mint 200 tudományos cikket írt a párhuzamos számítógépekről, párhuzamos programozási technológiáról és grid-rendszerekről. Számos európai grid-projektben (EDG, GridLab, SEEGRID, CoreGrid, GridCoord, ICEAGE, CancerGrid stb.) vett részt, mint a magyar tevékenység koordinátora. Jelenleg az európai FP7-es EDGI (European Desktop Grid Initiative) és SHIWA (Sharing Interoperable Workflows for large-scale scientific simulations on Available DCIs) projektek koordinátora. A Springer által kiadott Journal of Grid Computing folyóirat társ-főszerkesztője.

Irodalom

- [1] J. Cummings et al., Plasma Edge Kinetic-MHD Modeling in Tokamaks Using Kepler Workflow for Code Coupling, Data Management and Visualization, Com. Comp. Phys., Vol. 4(3), pp.675–702, Sept. 2008.
- [2] Ellis, C.A., Gibbs, S.J., Concurrency control in groupware systems. In: Proc. of the ACM SIGMOD'89 Conference on the Management of Data ACM, New York, pp.399–407.
- [3] MTA SZTAKI Grid Alkalmazás Támogató Központ (Grid Application Support Centre-GASuC), <http://www.lpds.sztaki.hu/gasuc/> (Letöltés: 2010. április 30.)
- [4] Dimitrios Georgakopoulos, Mark Hornick, Amit Sheth, An overview of workflow management: From process modeling to workflow automation infrastructure, Journal of Distributed and Parallel Databases, Vol. 3, Nr. 2, pp.119–153, April 1995.
- [5] Jonathan Grudin, Computer-Supported Cooperative Work: History and Focus. Computer 27 (5): 19–26, 1994.
- [6] Gergely Sipos, Péter Kacsuk, Multi-Grid, Multi-User Workflows in the P-GRADE Portal. Journal of Grid Computing, Vol. 3, Issue 3-4, Kluwer Academic Publisher, pp.221–238, 2006.
- [7] Gergely Sipos, Péter Kacsuk, Maintaining Consistency Properties of Grid Workflows in Collaborative Editing Systems, In: Proc. of 8th International Conference on Grid and Cooperative Computing (GCC09), IEEE Publishing, Lanzhou, China, pp.168–175, 2009.
- [8] Ian J. Taylor, Ewa Deelman, Dennis B. Gannon, Matthew Shields, Workflows for E-Science: Scientific Workflows for Grids, Springer-Verlag New York, Inc., Secaucus, USA (2006).
- [9] Workflow szócikk a Wikipedia-ban: <http://en.wikipedia.org/wiki/Workflow> (Letöltés: 2010. április 30.)