

AVTONOMNA PROGRAMSKA OPREMA IN PONOVDNA UPORABA OBSTOJEČIH INFORMACIJ

Dejan Lavbič

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Tržaška 25, 1000 Ljubljana
Dejan.Lavbic@fri.uni-lj.si

Povzetek

Večina računalniških programov je reaktivnih, saj se le odzivajo na vhodne attribute, ki mu jih podajo uporabniki. Ker lahko takšne sisteme uporabljamo zgolj kot orodja pri opravljanju nalog, v prispevku opisujemo agente, katerih glavna lastnost je avtonomnost, saj poleg reaktivnega poznajo tudi proaktivno, ciljno usmerjeno, delovanje. Vlogo takšnih sistemov vidimo predvsem v obliki pomočnikov, ki neodvisno opravljajo svoje delo s pomočjo metod umetne inteligence. Za demonstracija te ideje predstavljamo sistem i-novice za iskanje novic po elektronskih virih.

Abstract

AUTONOMOUS SOFTWARE AND REUSE OF EXISTING INFORMATION

Most of the software is reactive where it only responds to the input that users provide. Because we can use systems like that only as tools at performing tasks, we describe agents in this article with their main characteristic – autonomy, where beside reactive activity they also act proactive – goal oriented. We see the role of autonomous software as assistants that perform their tasks using artificial intelligence. Demonstration of this idea is presented in the system i-novice for searching the news on electronic sources.

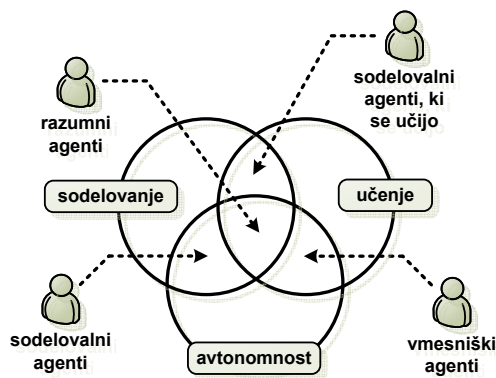
1 UVOD

Avtonomnost je lastnost, ki predstavlja zmožnost kontroliranja svojih lastnih dejanj. Ne samo, da lahko program prejme neposredne ukaze od uporabnika, ampak lahko konča opravilo brez posredovanja uporabnikov ali ostalih programov.

Ker obstaja splošno prepričanje, da je avtonomija ključna lastnost agenta, bomo v prispevku predstavili predvsem agente oz. inteligentne agente. Agenti se uporabljajo na številnih področjih, tako da trenutno ne obstaja univerzalna definicija pojma agent, vendar lahko kljub temu podamo definicijo, ki jo je v [5] zapisal M. Wooldridge: »**Agent je računalniški sistem, ki je nameščen v okolje in je sposoben avtonomnih akcij v tem okolju z namenom, da doseže načrtovane cilje.**«

1.1 Topologija agentov

Nwana [3] je klasificiral agente glede na tri osnovne attribute, ki naj bi jih agenti imeli: avtonomnost, sodelovanje in učenje. **Avtonomnost** je, kot smo že omenili, osnovna lastnost agenta, ki pomeni samostojno delovanje brez potrebne pomoči uporabnika. Namesto proaktivnega delovanja v okolju sami prevzamejo iniciativo [5]. **Sodelovanje** je zmožnost vzajemnega delovanja z ostalimi agenti in človekom s pomočjo ustreznega komunikacijskega jezika, kar že prehaja do socialno družbenih lastnosti. **Učenje** pomeni zmožnost izboljševanja učinkovitosti skozi čas.



Slika 1: Topologija agentov

Agente lahko poimenujemo po opravilih, ki jih opravljajo, zelo zanimiva pa je tudi klasifikacija glede na lastnosti (glej Slika 1), saj lahko na enostaven način določimo osnovne kategorije – sodelovalni, vmesniški, sodelovalni, ki se učijo in razumni agenti.

1.2 Sistem i-novice

Za namen predstavitve načrtovanja prilagodljivih in enostavnih metod za avtonomno izločevanje informacij iz spletnih virov smo pripravili spletni iskalnik i-novice, ki s pomočjo definiranih pravil oz. klasifikatorjev izločuje znanje s HTML strani. Pravila različnih spletnih virov se sistem samodejno nauči, tako da je v primeru sprememb na spletnih straneh (npr. spremeni se izgled portala 24ur.com) potrebno le ponovno izvesti fazo učenja, kjer se pravila v klasifikatorjih ažurirajo, tako da spletni iskalnik deluje pravilno. Uporaba pravil v obliki regularnih izrazov prinaša prilagodljivost avtonomnih informacijskih agentov za ponovno uporabo informacij, ki jih s pridom uporablja spletni iskalnik i-novice.

2 AVTONOMNI AGENTI

Avtonomno programsko opremo predstavljajo programi z lastnostjo agentov in vsaj v konceptualnem smislu predstavljajo stalno izvajajoč se proces. Agent lahko odkrije informacijo, ki je uporabniku potencialno zanimiva in se glede na to neodvisno odloči, da je pri tem dogodku potrebno obvestiti uporabnika. Agent lahko in praviloma ostane aktiven tudi po tem, ko uporabnik konča z vnosom podatkov oz. celo izključi svojo delovno postajo.

Zakaj potrebujemo avtonomne agente? Kot primer omenimo pomočnika v službi, ki mu je potrebno vsako opravilo natančno razložiti in demonstrirati, le ta verjetno ni preveč uporaben, saj potrebuje eksplicitna navodila in stalno nadzorovanje pri opravljanju dela. Glavna naloga pomočnika je predvsem razbremenitev nadrejenega, kjer naj asistent deluje čim bolj neodvisno. V računalništvu so nasprotje avtonomnosti programi, kjer se z uporabniškim vmesnikom komunicira s pomočjo ukazov oz. preko menijev. Za takšno vrsto, reaktivnih programov je značilno, da uporabnik najprej vnese vhodne podatke, nato glede na te parametre računalniški program izvede vrsto izračunov in prikaže rezultat uporabniku. Vsako dejanje, ki ga program izvede, mora uporabnik eksplicitno zahtevati. Avtonomni agent mora na drugi strani imeti zmožnost avtonomnega delovanja brez interakcije z uporabnikom. Kot primer omenimo program, ki uporabniku pošlje elektronsko sporočilo, ko pride do spremembe spletne strani, ki jo je uporabnik označil kot potencialno zanimivo.

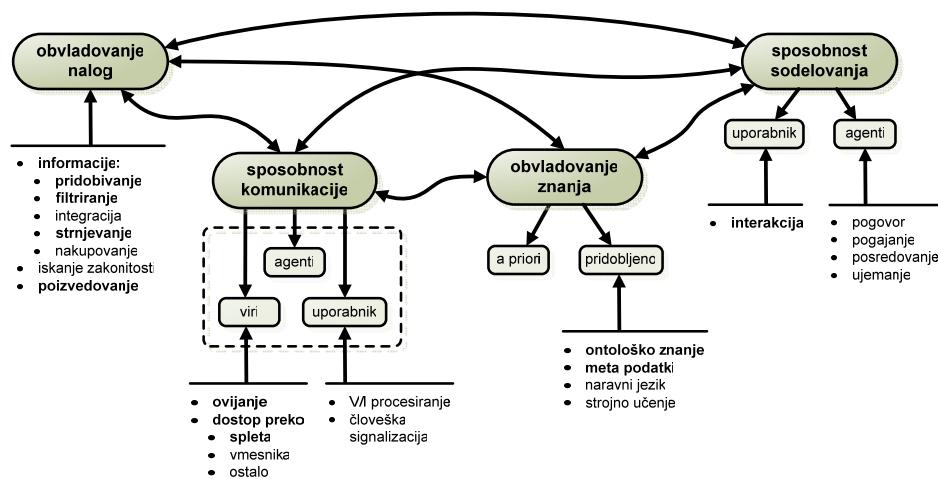
Veliko programskih rešitev za iskanje informacij na svetovnem spletu lahko poimenujemo agenti. Interakcija z uporabniškim vmesnikom je osnovno težava pri uporabi vseh trenutno

najbolj razširjenih spletnih iskalnikov (Google, Najdi.si, MSN, ...). Uporabniški vmesnik spletnega iskalnika je čisto običajen – uporabnik zastavi vprašanje in počaka na odgovor. Ko dobi odgovor, uporabnik nadaljuje s procesom brskanja po spletnih straneh, kjer obiskuje rezultate iskanja, ki mu ga iskalnik vrne, urejene po internem klasifikacijskem algoritmu, ki ga določen iskalnik uporablja. V naslednjem poglavju si bomo podrobneje ogledali agente, ki jih imenujemo informacijski inteligentni agenti in se uporabljajo predvsem za iskanje informacij po različnih porazdeljenih informacijskih virih in so pomembni za ponovno uporabo obstoječih informacij. Takšni agenti postajajo vedno bolj aktualni, saj količina informacij, ki je potencialno na voljo na svetovnem spletu, vključno s spletnimi stranmi, povezavami, ostalimi dostopnimi dokumenti in podatkovnimi bazami, še vedno hitro narašča. Trenutne statistike kažejo, da je bilo konec leta 2002 kar **320 milijonov uporabnikov interneta**, svetovni splet pa vsebuje več kot **800 milijonov spletnih strani**.¹ Statistični podatki o številu podatkovnih zbirk in številu podatkov, ki jih le-te vsebujejo, so zelo različni, medtem ko lahko le slepo ocenimo med koliko spletnimi stranmi in dokumenti (npr. PDF, ...) lahko iščemo po svetovnem spletu.

2.1 Informacijski inteligentni agenti

Informacijski agenti so posebna vrsta inteligentnih agentov, namenjeni obdelavi informacij iz množice porazdeljenih virov na internetu. Klusch [4] podaja definicijo informacijskega agenta: »**Informacijski agent je inteligentni agent, ki ima dostop do množice raznovrstnih, oddaljenih informacijskih virov in v imenu uporabnika ali drugih agentov namensko zbira, posreduje in vzdržuje informacije.**«

Slika 2 prikazuje sposobnosti informacijskih agentov, v povezavi s ključnimi tehnologijami, ki so navedene v referenci s sposobnostmi. Poudarjene so tiste značilnosti in tehnologije, ki so prišle do izraza pri implementaciji sistema i-novice, ki ga podrobneje opišemo v poglavju 2.3.



Slika 2: Osnovne sposobnosti informacijskih agentov

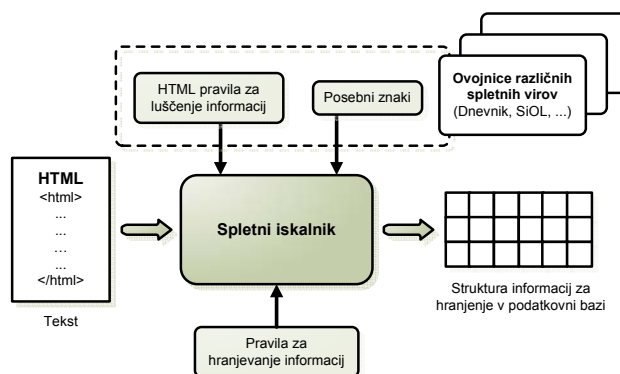
Sposobnost komunikacije informacijskega agenta vključuje komunikacijo z informacijskimi sistemi in podatkovnimi bazami, človeškimi uporabniki ali drugimi agenti. V komunikaciji se uporabljata agentna komunikacijska jezika FIPA-ACL in KQML. Informacijskemu agentu predstavitev in procesiranje ontološkega znanja in metapodatkov, profilov in naravnega

¹ <http://www.why-not.com/company/stats.htm>

jezika, prevedba podatkovnih formatov in aplikacije strojnega učenja omogočajo pridobivanje in vzdrževanje znanja o agentu samem in njegovem okolju. Sposobnost sodelovanja z ostalimi agenti je lahko odvisna od ustreznih tehnik koordinacije, kot so posredovanje storitev, ujemanje in pogajanje.

2.2 Luščenje informacij iz različnih spletnih virov in ponovna uporaba

Svetovni splet predstavlja zanimivo in rastoče področje za različne raziskovalne dejavnosti, kot so agenti in večagentni sistemi, pridobivanje informacij, programsko inženirstvo itd. Kljub izrednemu uspehu interneta je lahko ravno uspešnost njegova največja nevarnost: izjemna rast shranjenih informacij prinaša s seboj veliko težav, ki otežujejo pripravo in vzdrževanje spletnih aplikacij. Pravzaprav se povečuje zanimanje za gradnjo sistemov, ki znajo uporabiti obstoječe informacije, ki se v porazdeljeni obliki nahajajo na svetovnem spletu. Da se lahko lotimo gradnje takšnih sistemov, moramo analizirati in rešiti več problemov, kot so pridobivanje, luščenje in ponovna uporaba shranjenih informacij. Luščenje informacij je zelo zapletena naloga, ker večina elektronskih virov, ki so priključeni na svetovni splet, ne omogoča dostopa do svojih informacij v standardizirani obliki. Za luščenje informacij je predlaganih več rešitev; veliko jih je povezanih s semantičnim spletom, ostale pa uporabljajo druge rešitve, temelječe na XML-u, s pomočjo katerih koherentno predstavijo informacije, shranjene na internetu. S takšnim pristopom bomo lahko v prihodnosti gradili robustne in porazdeljene aplikacije, vendar se moramo zavedati, da je v tem trenutku semantični splet šele v razvoju. Tako moramo, če želimo zgraditi aplikacijo, ki bo ponovno uporabljala informacije, uporabiti druge pristope, ki sistemu omogočajo luščenje informacij.



Slika 3: Pristop ovojnica pri iskanju informacij

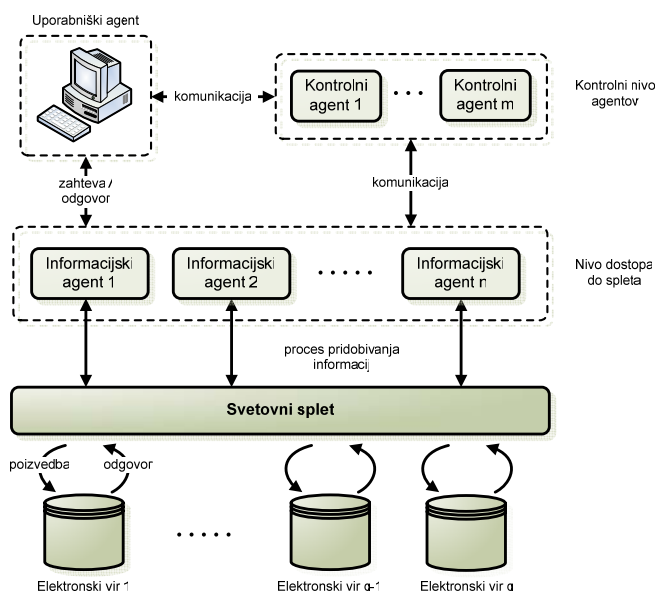
Pristop ovojnica (glej Slika 3) je eden najbolj uporabljenih. Ovojnice se uporabljajo za dostop do spleta kot relacijske podatkovne baze. Priprava teh ovojnica je lahko zelo zahtevna, predvsem takrat, ko pride do spremembe vira informacij (npr. spremeni se zunanja podoba oz. predstavitev informacij na spletnem mestu 24ur.com), saj moramo takrat ovojnice ponovno definirati. Za lažji razvoj in vzdrževanje ovojnica obstajajo številna orodja, kot so npr. W4F, WrapperBuilder itd.

2.3 Sistem i-novice

Sistem i-novic uporablja množico specializiranih agentov za pridobivanje novic iz določenih spletnih časopisov, ki imajo svojo vsebino javno objavljeno na svetovnem spletu. I-novice pridobivajo informacije iz izbranih elektronskih virov, filtrirajo različne odgovore specializiranih agentov in jih prikazujejo uporabniku.

Slika 4 prikazuje arhitekturo i-novic, ki jo lahko predstavimo v naslednjih medsebojno povezanih nivojih:

- **Uporabniški agent** predstavlja spletni uporabniški vmesnik, kjer lahko uporabniki iščejo po novicah, pregledujejo rezultate iskanja, nastavljajo svoj profil itd. Glavna prednost uporabniškega agenta je analiza vseh zahtev uporabnika in postopno prilagajanje uporabniku za večjo učinkovitost.
- **Kontrolni nivo agentov.** Ta del sistema sprejema zahteve s strani uporabnika in se odloča, ali je potrebno zahtevo poslati naprej v obdelavo spletnim informacijskim agentom, ki iščejo informacije na spletu, ali pa se potrebni rezultati že nahajajo v lokalnem repozitoriju. Glavna naloga je skrb za zagon spletnih informacijskih agentov, ki se lahko poženejo ročno, samodejno ali na zahtevo uporabniškega iskanja v uporabniškem vmesniku.
- **Nivo dostopa do spleta** predstavlja množico specializiranih spletnih informacijskih agentov, katerih primarna naloga je pridobivanje informacij iz različnih spletnih virov na internetu.



Slika 4: Arhitektura i-novic

3 METODE ZAGOTAVLJANJA AVTONOMNOSTI

Pri povečevanju avtonomije programske opreme je potrebno integrirati inteligenco oz. je potrebno razviti človeku podobne inteligentne agente z uporabo metod umetne inteligence, kot so:

- **Sklepanje na podlagi primerov²** je ena pomembnejših paradig sklepanja, kjer nove probleme rešujemo tako, da iz množice primerov poiščemo podoben primer, katerega rešitev že imamo.

² CBR – Case-Based Reasoning

- **Simbolično strojno učenje**³. Cilj SML pa je sklepanje novega znanja iz obstoječih podatkov za uporabo v prihodnje oz. sestavljanja znanja predvsem zaradi izboljšanja točnosti in učinkovitosti obstoječega znanja.
- **Sklepanje na podlagi pravil.**

Inteligenca je vgrajena programska komponenta, ki je sestavljena iz programske tehnologije in zapletenih algoritmov, kjer algoritmi vodijo programsko opremo pri izvajanju določenih opravil za uporabnika.

3.1 Odkrivanje zakonitosti v podatkih⁴

Iskanje zakonitosti v podatkih je področje s katerim se ukvarja veliko raziskovalnih skupin v umetnih inteligenci. Uporablja se na številnih drugih področjih, kot je odkrivanje znanja, OLAP sistemi itd. DM v podatkih vključuje luščenje ali odkrivanje uporabnega znanja oz. vzorcev iz velike količine podatkov, kjer se le ti lahko nahajajo v podatkovnih bazah, podatkovnih skladiščih ali ostalih informacijskih repozitorijih. DM lahko izvaja tudi čiščenje in integracijo podatkov, ki vsebujejo šum ali napake ter jih lahko pogosto uspešno preslika v uporabne podatke. Z dobro predstavitvijo podatkov lahko s pomočjo iskanja zakonitosti v podatkih zgradimo bolj točen model ali predstavitev znanja z uporabo različnih metod strojnega učenja, prav tako pa lahko izbere model z ovrednotenjem že naučenih modelov.

4 ZAKLJUČEK

Zmožnost agentov avtonomnega in vzporednega delovanja z uporabnikom postaja vse bolj pomembna lastnost pri interakciji človeka z računalnikom. Agenti bodo opazovali ljudi pri uporabi računalnikov oz. interakciji z uporabniškim vmesnikom in bodo uporabniku pomagali s prilagajanjem vmesnika in opravljanju določenih opravil namesto njega, medtem ko se lahko uporabnik osredotoči na reševanje drugih problemov. Tako lahko programska oprema predstavlja pomočnika, ne pa samo orodja za doseg določenih ciljev. Relativno slaba sprejetost agentov v širši javnosti je predvsem v pomanjkanju CASE orodij (in združljivost le teh s trenutnimi CASE orodji) ter splošno sprejetje ideje avtonomnih računalniških sistemov, kjer ljudje to področje še vedno obravnavajo z dvomom.

5 VIRI IN LITERATURA

- [1] LAVBIČ, Dejan: Uporaba inteligentnih agentov, diplomsko delo, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, avgust, 2004
- [2] LAVBIČ, Dejan: Povezava rezultatov iskanja spletnega inteligentnega agenta s podatki pomembnimi za poslovne odločitve, Zbornik posvetovanja DSI 2004, str. 670 - 675
- [3] NWANA, H. S.: Software agents: An overview, The Knowledge Engineering Review, 1996, št. 11(3), str. 1 – 40
- [4] KLUSCH, M.: Information Agent Technology for the Internet: A Survey, Journal on Data and Knowledge Engineering, Special Issue on Intelligent Information Integration, 2001, št. 36(3)
- [5] WOOLDRIDGE, Michael: An Introduction to MultiAgent Systems, John Wiley & Sons Ltd, 2002

³ SML – Symbolic Machine Learning

⁴ DM – DataMining