

MOŽNOSTI VYHLEDÁNÍ WEBOVÝCH SLUŽEB

WEB SERVICES DISCOVERY

Martin Smítka

Anotace:

Příspěvek se zabývá možnostmi nalezení webových služeb v prostředí Internetu zejména se zaměřením na necentrálně administrované prostředí.

Klíčová slova:

Komunikační protokol, XML, Webové služby, Komunikace

Abstract:

This thesis discusses web services discovery possibilities on the Internet from the machine computing perspective with the accent on the not centrally managed environment.

Key words:

Communication Protocol, XML, Web Services, Communication

ÚVOD

V současné době je možné na Internetu nalézt velké množství volně přístupných aplikací a nástrojů. Většinou jsou přístupné pomocí webových rozhraní samotných tvůrců, popřípadě jsou přístupné z tzv. mirror nebo alternativních odkazů.

Sémantický web naplňuje myšlenku přidat strojem zpracovatelné instrukce k datům a tím umožnit počítačům „rozumět“ obsaženým informacím a zpracovávat je namísto uživatele. Informace jsou určeny lidem, navíc je zde možnost automatizovaného zpracování. Webové služby posouvají využití webu ještě dále. Snaží se využívat webové prostředí jako infrastrukturu pro distribuované výpočty, integraci a automatizaci. FENSEL(6)

Nalezení vhodné informace/služby/souboru v prostředí webu je obecně náročnou operací ať již na určení konkrétního zdroje, či formátu. Webové služby dokáží zpřístupnit informace na webu i pro automatizované zpracování a posunout tak účinnost jejich vyhledávání a zpracování o úroveň dále. V tomto příspěvku jsou představeny možnosti komunikace s ohledem na možnost využití celkové infrastruktury v systému, který by umožňoval klientské části automaticky instalovat software.

CÍL

Cílem příspěvku je analyzovat komunikační procesy sloužící k nalezení webové služby a odvodit klíčové vazby a prvky, které bude možné použít pro vytvoření modelu vyhledávání webových služeb v systému správy SW v prostředí Internetu.

VÝSLEDKY

Pro nalezení konkrétní služby, která poskytne klientovi požadovanou odpověď je možné použít jeden z následujících obecných principů.

Centralizovaný přístup

V případě webových služeb jsou v současné době UDDI servery představiteli centralizovaného přístupu pro vyhledávání webových služeb v prostředí Internetu. UDDI servery slouží jako přístupový bod, ke kterému se případný zájemce o webové služby připojí a vyhledá požadovanou webovou službu. Přístup s využitím centralizovaného serveru s sebou nese několik závažných nedostatků týkajících se škálovatelnosti a přenesení zátěže vyhledávání poskytovatelů na stranu klienta. Taktéž centrální server představuje „single point of failure“, tedy kritický bod celého systému, neboť při jeho výpadku není možné použít žádnou službu.

○ Windows Server Update Services (WSUS)

Z pohledu konkrétního specializovaného aktualizacího software s centralizovaným přístupem (nevyužívajícím UDDI) lze na platformě Windows použít příklad WSUS, které slouží jako prostředek pro správu aktualizací pro celé síť. Základem je architektura klient – server. Cílem je udržení jednotné úrovně aktualizací operačního systému a pouze některých vybraných aplikací Microsoftu.

Umožňuje škálovatelnost pomocí kombinace několika WSUS serverů rozprostřených v rámci celé sítě. Tyto servery jsou zpravidla v hierarchii, tzn. existuje jeden, nebo několik serverů, které jsou klíčové pro chod celé podřízené hierarchie. WSUS je navržen pro prostředí, které je administrováno centrálně, popř. je administrace delegována.

Klientská část WSUS je konfigurovatelná pomocí skupinových politik (místní počítač či doménové politiky), popřípadě ve výchozím nastavení využívá Windows update. V případě domén nejsou klienti z pohledu administrace autonomními jednotkami.

Peer to Peer síť

Základem komunikace je princip sítě, která nemá centrální uzel. Není tedy možné znepřístupnit celou síť vyřazením několika klíčových serverů. V rámci sítě spolu účastníci vzájemně spolupracují. Klientská část aplikace odkazuje na veřejně přístupný seznam serverů, které umožňují propojení jednotlivých klientů a prohledávání sdíleného prostředí. Způsob spolupráce mezi účastníky sítě dále umožňuje členit varianty peer to peer sítě.

Klasický peer to peer

V tomto způsobu komunikace jsou všichni účastníci zároveň klienty i servery ostatním účastníkům. Každý účastník musí znát alespoň jednoho dalšího účastníka sítě, kterému adresuje své požadavky. Pokud tento účastník nezná odpověď na požadavek, přepoše požadavek dalším účastníkům.

Hybridní peer to peer

V případě hybridní sítě dochází ke kombinaci centralizovaného přístupu s peer to peer sítí. V síti existují dva typy účastníků: tzv. super uzly a uzly potomků. Každý uzel potomka zná právě jeden super uzel. Super uzel slouží jako registrátor uzlů potomků, jejichž seznam spravuje, a zároveň jako registrátor sousedících super uzlů.

○ Gnutella

Gnutella je protokol a zároveň i aplikace využívající tento protokol ke sdílení souborů v peer to peer síti. Každý účastník sítě slouží zároveň jako SERVER i cliENT (dále označován jako servent). Servent vyhledává ostatní servery díky ping-pong mechanismu. Východiskem je znalost serventa alespoň jednoho dalšího servanta, který je téměř stále v provozu. Dotaz na

vyhledání ostatních serventů ping je odeslán druhému serventu, který ping přepoše všem jemu známým servetům atd. Klíčový je zde parametr TTL (time to live), který omezuje dosah pingu. Odpověď pong obsahuje zejména údaje o nalezeném serventu, tedy údaje o konektivě (IP, port), počty a velikost sdíleného obsahu.

- Web Services Peer-to-peer Discovery Service (WSPDS)

BANAEI-KASHANI(8) a kol. představují model založený na protokolu Gnutella pro vyhledávání webových služeb a tím odstraňují závislost na určitém serveru. Každý servent poskytuje informace o vlastních webových službách a zároveň se dotazuje na služby, které nejsou v jeho seznamu, u ostatních serventů. V okamžiku, kdy servent obdrží žádost na službu, o které nemá informace uložené lokálně, v roli klienta zahájí vyhledávací proces přes své sousedy do sítě. Vyhledávací proces končí v okamžiku, kdy některý z dotazovaných serventů najde lokálně požadovanou službu a informace o ní vrátí serventu, který začal vyhledávání.

Peer to peer sítě mají bohužel některé nevýhody, které plynou z jejich podstaty. Příkladem může být zahlcování sítě, nemožnost garance úplné odpovědi. Účinnost vyhledávání ostatních uzlů lze optimalizovat.

Sdílený prostor

FENSEL(6) předkládá pohled na webové služby, který je odlišuje od webu v dnešní podobě. Zdůrazňuje základní odlišnosti plynoucí ze způsobu adresování zdrojů na webu a přístupu k zdrojům přístupných z webových služeb. Řešení některých potíží webových služeb je v těchto souvislostech spatřováno ve vytvoření sdíleného prostoru, do kterého by měly strojem řízené aplikace přístup (tuple space). Tento sdílený prostor pro všechny webové služby dokáže skrýt tři dimenze výměny informací:

- vztahy: strany, které si ve skutečnosti vyměňují informace nemají přímý kontakt s druhou stranou. Vztah je skryt uložením informace v neutrálním prostoru a následném vyzvednutí druhou stranou bez nutnosti znalosti, kdo informace uložil.
- čas: nezáleží na tom, zda je druhá strana v daném okamžiku přístupná.
- prostor: procesy mohou běžet v oddělených prostředích, ve kterých je pojitkem přístup do společného sdíleného prostoru.

Díky těmto třem dimenzím je tento sdílený prostor nazýván „Tripple space“.

Hledání informací v prostředí webu je dnes založeno na použití klíčových slov, frází, omezování podmínek. V okamžiku, kdy bereme prostředí webu jako společný prostor, je pro lidského uživatele reprezentace na webové stránce zcela zřejmá bez ohledu na použité značkování (XML, HTML), grafické styly (CSS), technologie zobrazení (HTML text, Flash, Obrázky). V případě strojově zpracovatelných instrukcí přibývá nezbytně nutný krok přeložit použité značkování (přidaná sémantický data) do slovníku podporovaného strojem, tedy zajistit, aby se stroj dokázal (nejlépe dynamicky) naučit a obsahově zařadit značkování, které zpracovává poprvé. V okamžiku, kdy by se prostor pro strojové zpracování řídil jedním standardem (jako vizuální reprezentace určená lidem), byl by problém vyřešen.

Pro strojově zpracovatelné vyhledávání je nutný předpoklad, že stroj ví, co hledá a jakou odpověď typově obdrží. Klasifikace zdrojů a jejich kontextu (přidání sémantických dat, RDF, OWL) bude zřejmě vždy ovlivněno pohledem tvůrce. Otázkou tak zůstává, zda „naučit“ stroj používat webové služby s třeba ne zcela vyhovujícími sémantickými daty, nebo vytvořit plně virtuální svět v podobě klasifikace, která zohlední všechny možné pohledy na daný zdroj. To by ve své podstatě pro stroj nakonec znamenalo dosažení úrovně lidského chápání zobrazené webové stránky.

ZÁVĚR

Prostředí Internetu je velmi dynamické a různorodé a orientace v něm je někdy velmi nesnadná. Pro administraci koncových (domácích) uživatelských stanic neexistuje obecně používaný model umožňující distribuci aplikací. Z pohledu koncových uživatelů a principu fungování Internetu se zdají být nejvhodnější technologie, které nevyžadují žádnou centrální autoritu a tomuto odpovídají zřejmě nejvíce peer to peer sítě.

Při tvorbě dnešních webových služeb lze využít již existujícího standardu značkování (třeba obecnějšího a odlišnosti v úhlu pohledu implementovat v logice vlastní aplikace) nebo vytvořit vlastní a vytvořit propojení na stávající standardy.

Je v zájmu výrobce software zpřístupnit své produkty veřejnosti, to se dnes děje díky možností stažení na vlastních webových stránkách. Pokud by výrobce použil rozhraní v podobě standardizované webové služby, bylo by možné kombinací těchto rozhraní v peer to peer síti vytvořit specializované prostředí s nabídkou stažitelného software. Výhody využití peer to peer sítě je možné pro tvůrce software spatřovat právě v autonomii každé služby.

Je možné předpokládat vytvoření takového modelu, kdy by za poskytování služeb v podobě přístupu koncových stanic k takovému systému bylo možné vybírat mikroplatby, které by byly opodstatněné filtrováním a hodnocením sdílených informací v síti.

LITERATURA

BUSSLER, CH.: *A Minimal Triple Space Computing Architecture*, [on-line], [cit. 2006-06-27], URL: <http://dip.semanticweb.org/documents/bussler-wiw05.pdf>

TOMA, I. – SAPKOTA, B. – SCICLUNA, J. a kol.: *A P2P Discovery mechanism for Web Service Execution Environment*, [on-line], [cit. 2006-06-27], URL: <http://www.dip.semanticweb.org/documents/p2pDiscovery.pdf>.

YANG, B. - GARCIA-MOLINA, H.: *Comparing Hybrid Peer-to-Peer Systems*, [on-line], [cit. 2006-06-27], URL: http://www.dia.uniroma3.it/~vldbproc/060_561.pdf

CHAWATHE, Y. – RATNASAMY, S. – BRESLAU, L. a kol. *Making Gnutella-like P2P Systems Scalable*, [on-line], [cit. 2006-06-27], URL: <http://www.acm.org/sigs/sigcomm/sigcomm2003/papers/p407-chawathe.pdf>

The Annotated Gnutella Protocol Specification v0.4, [on-line], [cit. 2006-06-27], URL: <http://rfc-gnutella.sourceforge.net/developer/stable/index.html>

FENSEL, D.: *Triple-space computing: Semantic Web Services based on persistent publication of information*, [on-line], [cit. 2006-06-27], URL: <http://dip.semanticweb.org/documents/Triple-basedComputing-SWS2004.pdf>

PAOLUCCI, M. - SYCARA, K. - NISHIMURA, T. a kol.: *Using DAML-S for P2P Discovery*, [on-line], [cit. 2006-06-27], URL: http://www.cs.cmu.edu/People/softagents/papers/p2p_icws.pdf

BANAEL-KASHANI, F. - CHEN, CH. – SHAHABI, C.: *WSPDS: Web Services Peer-to-peer Discovery Service*, [on-line], [cit. 2006-06-27], URL: http://infolab.usc.edu/DocsDemos/isws2004_WSPDS.pdf

KONTAKT

Ing. Martin Smítka

Česká zemědělská univerzita v Praze - Katedra informačních technologií

Provozně-ekonomická fakulta, Kamýcká, 165 21 Praha 6 – Suchbátka

732657308

krysolet@pef.czu.cz