

Bevezető

A *szemantikus világháló* (angolul: *Semantic Web*) a számítástudomány egy új kutatási-fejlesztési iránya, amelynek fő célja, hogy a világhálón elérhető témérdek információt számítógépes feldolgozásra alkalmasabbá tegye. Az irányzat jelszava, hogy a világhálón található információt a számítógépek *ne csak olvasni, hanem értelmezni* is tudják. Ehhez szükséges, hogy a hálón levő adatokhoz ún. metainformációt társítsunk, például egy képhez adjuk meg, hogy ki készítette, mi a címe, mit ábrázol stb. Ezen túlmenően a számítógépeket képessé kell tennünk következtetések elvégzésére, például ha egy képről ismert, hogy egy folyót ábrázol, akkor ebből gépi úton ki lehessen következtetni azt is, hogy víz látható rajta.

A szemantikus világháló irányzatához szorosan kapcsolódik az ontológiák, azaz hierarchikus fogalmi rendszerek kutatása és fejlesztése. Az ontológiákon való következtetések alapját pedig az ún. leíró logikai (*Description Logic*) formalizmusok képezik. Nem véletlen, hogy a World Wide Web konzorciumnak (W3C) a szemantikus világhálóhoz kötődő legújabb szabványa, az OWL nyelv, már leíró logikai alapokon nyugszik, a nyelv egyik szerzője a leíró logikai következtetési rendszerek nemzetközileg elismert szakértője.

Az ontológiai rendszerek és a metainformációk ugyanakkor egyre nagyobb szerepet kapnak a világhálón kívüli információforrások, például a vállalati adat- és tudásbázisok kezelésében is. Az elérhető információ mennyisége – mind a világhálón, mind azon kívül – óriási ütemben nő, emiatt egyre nagyobb szükség van az adatforrások rendszerezésére, összekapcsolására. Ebben is segítenek a könyvben bemutatott új eszközök és módszerek.

Kinek szól ez a könyv?

A könyvet elsősorban felsőoktatási tankönyvnek szánjuk, az egyre több intézményben megjelenő, a szemantikus világhálóval és a hozzá kapcsolódó témakörökkel foglalkozó kurzusokhoz. A könyv kidolgozásához felhasználtuk a BME Villamosmérnöki és Informatikai Karán a szerzők által tartott, „A szemantikus világháló és az ontológiakezelés alapjai” című tárgy előadásanyagát és oktatási tapasztalatait.

Emellett a könyvet ajánljuk minden olyan érdeklődőnek is (legyen az szakmabeli vagy sem), akit megérintett valaha az Internet szelleme, elgondolkodott már azon, hogyan működnek a keresőrendszerek, hogyan épül fel a web, milyen lehetőségek rejlenek benne. Ajánl-

juk továbbá mindazon matematikai beállítottságú olvasóknak, akiket érdekelnek a tudás-reprezentációs formalizmusok, a leíró logikák, és az ezekhez kidolgozott gépi következtetési módszerek.

Végül ajánljuk a könyvet a programozás kedvelőinek, mert a különböző problémákhoz kapcsolódóan számos esetben közlünk megvalósítható algoritmusokat és optimalizálásokat, illetve mutatunk be implementációs részleteket.

A könyv felépítése

Könyvünk célja, hogy a szemantikus világháló területének gyakorlati és elméleti oldalát egyaránt bemutassa. Ennek megfelelően a könyv három nagy részre oszlik. Az első rész ismerteti a szemantikus világháló irányzatát, a második rész tárgyalja az ontológiákat és speciálisan a leíró logikán alapuló rendszereket, végül az utolsó rész foglalkozik az ontológiák tágabb felhasználási lehetőségeivel, valamint a komplex ontológiakezelő, ontológia alapú tudásreprezentációs rendszerekkel.

Az alábbiakban megadjuk a könyv egyes fejezeteinek rövid leírását.

I. rész – A szemantikus világháló

Első fejezet – A világháló napjainkban

Az első fejezet célja megismertetni az olvasót az internetes keresőrendszerek működési elveivel. Bemutatjuk az intelligens keresést nehezítő főbb tényezőket, a gépi úton fel nem dolgozható információkat, a keresőcsapdákat és a szemantika hiányából eredő problémákat. Ismertetjük a jelenleg elterjedt technológiák által kínált megoldási lehetőségeket. A megértést segítően a fejezet elején összefoglaljuk az Internetről szóló alapvető ismereteket, tisztázzuk a statikus és dinamikus weboldalak fogalmát. A fejezethez kapcsolódóan a függelékben kiegészítő anyagként néhány elterjedt webes technológiát mutatunk be.

Második fejezet – A szemantikus világháló és az RDF elképzelés

A második fejezetben ismertetjük a szemantikus világháló irányzatát, amely arra az alapötletre épít, hogy társítsunk metainformációkat internetes erőforrásokhoz, valamint következtesünk azokon. A fejezet elején felvázoljuk, hogy ez az ötlet hogyan segíthet megoldani mindazon problémákat, amelyekről az első fejezetben írtunk. A fejezetben részletesen ismertetjük a szemantikus világhálót megalapozó nyelveket: az RDF-et és az RDF sémát. A fejezetet néhány RDF-esettanulmánnyal zárjuk.

Harmadik fejezet – RDF-források feldolgozása és lekérdezése

A harmadik fejezetben bemutatjuk az RDF formátumú metaadatok tárolásának és lekérdezésének különböző módjait. Ismertetünk néhány létező XML-, illetve RDF-elemzőt és lekérdezőnyelvet, valamint érvelünk amellett, hogy miért nem alkalmasak a közönséges XML-lekérdezők RDF-adatok kezelésére. Megmutatjuk, hogy milyen fajta következtetési feladatok merülnek fel a lekérdezések során. A fejezet zárásaként bemutatjuk, hogy milyen optimalizációs lehetőségeink vannak, amelyekkel hatékonyabbá tehetjük az RDF-lekérdezések végrehajtását.

II. rész – Ontológiák

Negyedik fejezet – A leíró logikák

A fejezetben bemutatjuk a leíró nyelveken alapuló tudásreprezentációs rendszerek felépítését, bevezetjük az ún. terminológiai tudást tároló T-doboz, illetve a konkrét adatokról szóló A-doboz fogalmát. Ezután a leíró logikai nyelvcsalád legfontosabb tagjait ismertetjük, a legegyszerűbb \mathcal{AL} nyelvtől, a közepes bonyolultságú \mathcal{ALCN} nyelven keresztül egészen a \mathcal{SHIQ} rövidítéssel jelölt fejlett leíró logikai nyelvig. Foglalkozunk a leíró logikákhoz kapcsolódó következtetési feladatok meghatározásával és rendszerezésével, valamint a leíró logika és az elsőrendű matematikai logika közötti összefüggésekkel is. Végül röviden áttekintjük a \mathcal{SHIQ} nyelven is túlmutató fejlett nyelvi elemeket.

Ötödik fejezet – Következtetés leíró logikákon

A fejezet konkrét következtetési algoritmusokat ismertet. Bemutatja az egyszerűbb leíró logikai nyelvekre alkalmazható strukturális tartalmazási algoritmust, majd a bővebb nyelveket is kezelni képes tabló-algoritmus \mathcal{ALCN} , illetve \mathcal{SHIQ} változatát. Az ismertetett módszereket bőséges példaanyaggal illusztráljuk, az algoritmusok legfontosabb tulajdonságait matematikai pontossággal bizonyítjuk. A fejezet zárószakaszában a \mathcal{SHIQ} tabló-algoritmus optimalizálásával foglalkozunk.

Hatodik fejezet – Egy leíró logikai következtető megvalósítása

Ebben a fejezetben részletesen ismertetjük a \mathcal{ALCN} tabló algoritmusnak a Haskell funkcionális programozási nyelven való megvalósítását. Ezzel az a célunk, hogy egy közepes bonyolultságú leíró logikai következtető motor tömör, könnyen átlátható és jól olvasható „referencia-implementációját” mutassuk be. A Haskell nyelv ismeretét nem tételezzük fel, a szükséges nyelvi elemeket megmagyarázzuk. Ez a program – amellett, hogy ténylegesen futtatható – a matematikai jelöléshez nagyon közel álló formalizmust használ, ezért úgy gondoljuk, hogy azok számára is könnyen érthető, akik a funkcionális programozási nyelvekkel még nem foglalkoztak.

III. rész – Ontológiák felhasználási lehetőségei

Hetedik fejezet – Ontológiák és a szemantikus világháló

A fejezetben bemutatjuk az OWL ontológianyelvet, amely a leíró logikákon alapszik, és amelyet az RDF séma nyelvének kibővítéseként alkottak meg. Részletesen ismertetjük az egyes nyelvi konstrukciókat, valamint megadjuk azok leíró logikai megfelelőit. A fejezet második részében megmutatjuk, hogy az RDF séma valójában több, mint aminek látszik. Ehhez ismertetjük az RDF séma modellelméletét és összevetjük azt a leíró logikák és más logikák modellelméletével. A fejezet végén röviden bemutatjuk a szemantikus webszolgáltatásokat, amelyek OWL nyelvű ontológiákat használnak az összetettebb szolgáltatások leírásához.

Nyolcadik fejezet – A SILK rendszer

A könyv zárófejezete bemutatja, hogy milyen feladatai vannak egy összetett ontológiakezelő rendszernek, valamint részletesen ismertet is egy ilyet, a SILK rendszert, amely egy sikeres európai uniós projekt eredménye. A SILK célja heterogén vállalati információforrások

szintaktikus és szemantikus integrációja. Ismertetjük a SILK rendszer azon komponenseit, amelyek az ontológiák összehasonlítását, verifikációját és komplex lekérdezésekben történő felhasználását támogatják. A fejezetben bemutatjuk a megvalósítás néhány érdekes részletét is.

Kiegészítő anyagok

A könyv anyagának kiegészítéseként a kiadó és a szerzők honlapján egyaránt közzéteszünk fontos anyagokat: a program példák forráskódját, a nyelvek szintaktikus leírását, valamint további kiegészítő anyagokat [63].

A könyv szerzői

A könyv 1–3. fejezete, a 7. fejezet, valamint a függelék Lukácsy Gergely munkája. A 4. fejezetet, valamint az 5.1–5.3. szakaszokat Szeredi Péter írta. Az 5.4. szakasz szerzője Nagy Zsolt. A 6. fejezet Benkő Tamás és Nagy Zsolt közös munkája. A 8. fejezetet Benkő Tamás írta. A könyv szakmai szerkesztője Szeredi Péter.

Hogyan olvassuk a könyvet?

A könyv szerzői természetesen azt javasolják, hogy az olvasó a fejezeteket a könyvbeli sorrendben olvassa végig. Azonban szeretnénk segíteni azoknak az olvasóknak, akik már rendelkeznek bizonyos előismeretekkel, vagy a könyv témakörén belül csak egyes részterületek iránt érdeklődnek.

Az egyes fejezetek kisebb-nagyobb mértékben támaszkodnak az előző fejezet(ek)ben elmondott ismeretekre. Ezeket a függőségeket a túldoldali ábrán tüntettük fel. A csomópontok a fejezeteket reprezentálják (a szöveg a fejezet tartalmára utal, a szám annak sorszámát adja meg), a pontozott vonalak az ajánlott belépési pontokat jelentik.

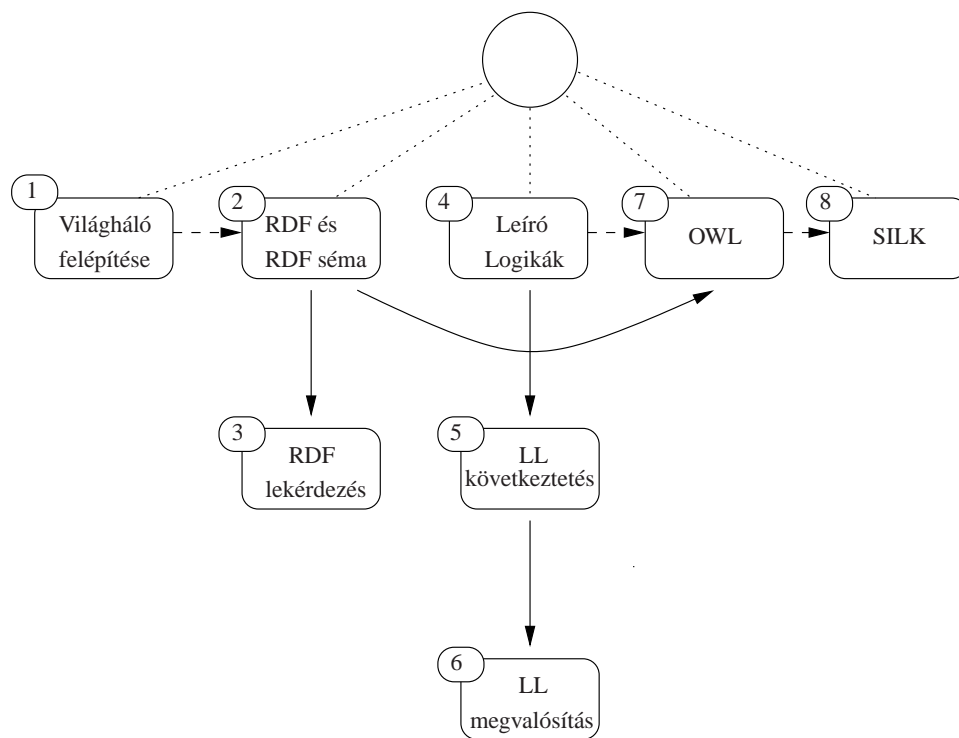
Folytonos irányított vonallal jelöltük az *erős*-, szaggatottal a *gyenge* függőségeket. Az előbbi esetben úgy érezzük, hogy a függőségi gráf követése elengedhetetlen a megértés szempontjából, az utóbbi esetben „csupán” nagymértékben elősegíti azt.

Erős függőségre példa a negyedik, ötödik és hatodik, illetve a második, harmadik és hetedik fejezetek láncolata. Nagyon nehezen lehetne egy leíró logikai következtető algoritmus működésének végiggondolását a leíró logikai alapismeretek hiánya, vagy az RDF-alapú OWL nyelv megértését az, ha nem ismernénk az RDF nyelvet.

Gyenge függőségre példa a hetedik és negyedik fejezet kapcsolata. Az előbbiben ismertetett OWL konstrukciókhoz megadjuk (a negyedik fejezetbeli) leíró logikai megfelelőjüket is, ugyanakkor reményeink szerint a hetedik fejezet ezek ismerete nélkül is feldolgozható.

Köszönetnyilvánítás

A könyv szerzői szeretnének köszönetet mondani mindazoknak, akik segítettek a könyv létrejöttét. Köszönet illeti a szakmai lektorokat, Varró Dánielt és Varró-Gyapay Szilviát gondos munkájukért, értékes megjegyzéseikért. A könyv egyes fejezeteinek érthetőbbé tételében, a hibák kijavításában sokat segítettek Laczay Bálint és Szabó Péter észrevételei. Sokat köszönhetünk a BME Villamosmérnöki és Informatikai Karán tartott „A szemantikus világháló és



az ontológiakezelés alapjai” című tárgy hallgatóinak, akiken a könyv anyagát kipróbálhattuk és reakciójukból sokat tanulhattunk.

A 8. fejezet a SILK rendszert mutatja be, amely egy nemzetközi csapat több évi munkájának eredménye, és amelyet a jelen könyv szerzőinek részvételével az IQSOFT Rt. illetve IQSYS Rt. cégeknél dolgoztak ki, francia, román és görög partnerekkel együtt. Szeretnénk tehát köszönetet mondani a SILK projekt összes résztvevőjének, és a magyar projektcsapat minden tagjának név szerint is: Fokt Attilának, Kilián Imrének, Kósa Mártonnak, Krauth Péternek, aki az egész nemzetközi projekt vezetője volt és Ruttnerné Kassai Dórának.

Köszönettel tartozunk az Oktatási Minisztériumnak a *Felsőoktatási Tankönyv- és Szakkönyvtámogatási Pályázat* keretében nyújtott támogatásért, valamint a Typotex kiadó minden érintett munkatársának a könyv megjelentetésében nyújtott segítségükért.