



Beszámolók ◦ Szemlék ◦ Közlemények

INFORMÁCIÓS RENDSZEREK

CONDOR: integrált információs rendszer
a Siemens fejlesztésében

A CONDOR kutatási projekt

A CONDOR (*Communication in Natürlicher Sprache mit Dialog-Orientierten Retrievalsystem = információközlés természetes nyelven párbeszédés keresőrendszerrel*) rendszer fejlesztésének célja a 80-as évekre olyan integrált rendszer kialakítása, amely mind strukturált (adatszerű), mind nem-strukturált (szövegszerű) információk kezelését lehetővé teszi. A rendszert a Siemens AG fejleszti 1973 óta a német szövetségi kutatási és technológiai minisztérium 50%-os pénzügyi támogatásával.

A jelenlegi információs rendszerek

A jelenlegi gépi információs rendszerek hátránya, hogy a felhasználói kényelmet e rendszerek tervezői nem vették tekintetbe, így a felhasználók (elsősorban a közvetlen felhasználók) kelletlenül alkalmazzák azokat. A felhasználó és a rendszerek közé ékelődnek szükség-szerűen a rendszerszakértők. A CONDOR rendszer tervezői ki akarják küszöbölni az ember és az információs rendszer közötti, használatot gátló tényezőket.

A strukturált – ismert és mindig változatlan, egységes formátumú, szerkezetű és szervezésű – információkat kezelő adatbank-rendszerek jöttek létre először, és ezeket követte az elsősorban a szakirodalmi dokumentumokat (szövegeket, nem-strukturált információkat) feldolgozó rendszerek megjelenése.

Ez utóbbi rendszerekre nézve – munkamódjuktól szinte függetlenül – az alábbi közös jellemzők emelhetők ki:

szövegeket (dokumentumokat) tárolnak;

a tárolt szövegekre mutató *deskriptorokkal* (kulcszavakkal) dolgoznak;

a tárolt szövegek keresése esetleg *párbeszédés üzemmódban* is lehetséges;

a keresés a deskriptorokra nézve szigorúan *formalizált utasítások* bevitelével kezdődik;

a fellelt dokumentumok kiadása szintén megfelelő *utasítások* hatására történik meg.

A felhasználói kényelem

A felhasználás kényelmére nézve mértékadó az információs rendszer kezelhetősége és az információs rendszer által végrehajtható funkciók készlete.

A rendszernek olyannak kell lennie, hogy abban *bárki könnyen végezhesen keresést*, mert ezt optimálisan csak az képes elvégezni, akinek az információra szüksége van. Ritkán tudják ugyanis pontosan megfogalmazni a szükségelt információt (esetleg nincsenek is pontosan tudatában a tartalmi vonatkozásoknak), és csak a rendszerrel való párbeszéd folyamán sikerül a kérdést körülírni, meghatározniuk.

Az információs rendszernek egyidejűleg kell képesnek lennie *numerikus és szöveges*, azaz strukturált és nem-strukturált információkkal folytatott munkára, azaz mind az adatbankok, mind a szöveges adattárak kezelési funkcióinak ellátására.

Ezekből a szempontokból vizsgálódva a jelenlegi információs rendszerek hiányosságokat mutatnak. A nem szakember rendszerint nem képes kezelni e rendszereket és az integrált funkciók ellátása sincs biztosítva bennük. A létező rendszerek sokfélesége okozza azt, hogy az információk gépi kereséséhez szükséges ismereteket (elsősorban az alkalmazott osztályozási rendszerek, programozási nyelvek és kezelési utasítások ismeretét) még az információs szakember számára is igen nehéz megszerezni.

Ezenkívül az információk *szellemi munkával* (indexelés, tárgyszavazás) történő vagy *gépi* (automatikus indexelés pozitív vagy negatív szótárazással, illetve a számítógépen tárolt fezaurusz felhasználásával) *tartalmi feldolgozása* is problémákat rejt magában. Sőt, így eljárva a felhasználónak előbb rá kell bukkannia a részére (esetleg még nem is teljesen felismerten) szükséges információt jellemző tárgyszóra vagy deskriptorra és csak azután láthat hozzá magának az információnak a kereséséhez.

A CONDOR projekt célkitűzése

A CONDOR projekt alapvető célkitűzése az optimális felhasználói kényelmet biztosító és nagy funkciókészlettel rendelkező programcsomag előállítás. A funkcionális tervezés során nemcsak az adatbank és a szöveges adattár integrált kezelésével kapcsolatos feladatok megoldását biztosították, hanem a rendszer ún. *mesterséges intelligenciáját* is (természetes nyelvű szövegek automatikus elemzése, szemantikai hálók feldolgozása, tanuló- és osztályozási algoritmusok stb.).

A kereső kérdés természetes nyelven is feladható a rendszernek. Pontos megfogalmazása a rendszer által kiadott űrlapok (a display terminálon jelennek meg) kitöltésével, azaz a válaszok bebillentyűzésével történhet meg.

A CONDOR összetevő elemei

A rendszerhez először *két részrendszert* – a strukturált és a nem-strukturált adatokat kezelő részrendszereket – fejlesztettek ki és ezek kipróbálása után kerül sor összekapcsolásukra az előírt funkciók biztosítása, a két részrendszer integrálása érdekében.

A strukturált információkat kezelő részrendszerként született meg az *SPB (Strategische Problem-Beschreibung = problémaleírési stratégia)*, amely elem a kérdés pontos megfogalmazását teszi lehetővé, valamint módot ad az információk párbeszédés átszerkesztésére, változtatására. Alkalmazásához azonban a felhasználónak nincs szüksége az adatokat leíró nyelv ismeretére.

A rendszer az alábbi *információfeldolgozási műveleteket* végzi el a felhasználói oldalon:

az adatok begyűjtése, plauzibilitás és konzisztencia-vizsgálat;

keresés súlyozható keresési kritériumokkal, szemantikai hálóban, fonetizálva és a felhasználó által kitöltendő űrlapokat alkalmazó keresőnyelvvél;

szerkesztett formában történő adatkiadás;

az elvégzett műveletek jegyzőkönyvezése;

a tárolt információknak szabadon választható szelekciós kritériumok szerinti értékelése.

A *szöveges információkat kezelő részrendszer* elemei a közvetlen optikai karakterolvasásos bevitel, a természetes nyelvű szövegek automatikus elemzése, automatikus indexelés, illetve osztályozás és az így leírt dokumentum-állományban végezhető – a rendszer által vezérelt – automatikus keresés.

A cél a szövegelemző komponens fejlesztése volt, azzal a feltétellel, hogy mind a szövegek leírásánál, mind keresésénél egységesen és automatikusan alkalmazható legyen. Ezzel kiküszöbölhető a jelenlegi rendszereknek az a hátránya, hogy különböző leírási és kérdezési rendszerek alkalmazása miatt bonyolultak és nehezen kezelhetők. A rendszer az alkalmazott eljárás révén maga veszi át a *feldolgozás feladatát*, a tárolt információkban – a rendszer rendelkezésére álló saját ismeretei alapján – a *keresést automatikusan végzi el* és az eredményt a kérdéssel való megfelelés aránya szerinti rangsorba rendezve szolgáltatja.

Ha a rendszer maga nem jut kielégítő eredményre, akkor párbeszédés módon fordul a felhasználóhoz a kérdés kiegészítésére.

A jövő és a próbaüzem

A két részrendszert modell formájában már kipróbálták. Három próbaüzemet terveznek:

COBIS (Computer-Orientiertes Büroinformations-system = számítógépes hivatali információs rendszer) az elnevezésének megfelelő feladatokkal a regensburgi egyetemen;

GIZ (Giftinformationszentrale = Mérgezési-Információs Központ) együttműködésben a Mérgezési Tanáccsal és az Orvosi Információs Központtal;

ASIS (Arbeitschutz-Informationssystem = Munkavédelmi Információs Rendszer) együttműködésben a dortmundi Munkavédelmi és Balesetkutatási Szövetségi Hivattal.

/TAEUBER, D.: CONDOR: Ein integriertes Datenbank- und Informationssystem = Nachrichten für Dokumentation, 29.köt. 3.sz. 1978. p.127–130./

(Schiff Ervin)



Az ipari információ, mint a nemzeti információs rendszer alkotóeleme

Az iparvállalati információ általában nem tekinthető a nemzeti információs rendszer részének, mivel csupán egyes részei (pl. tudományos kiadványok, szabadalmak) hozzáférhetőek mindenki számára, más elemei azonban nem, vagy csak nagyon korlátozottan.