

Búzanemesítési információs rendszer

KUTI CSABA – LÁNG LÁSZLÓ – BEDŐ ZOLTÁN

MTA Mezőgazdasági Kutatóintézete,
Martonvásár

Összefoglaló

Az elektronikus vezérlésű parcellavetőgépek korában akár napi több tízezer parcella vetése sem lehetetlen. Jelentős nehézséget okoz viszont a parcellákhoz kapcsolódó összes adat kezelése. A legnehezebb feladatot mégis az jelenti, hogy hasznos, értelmes információt nyerjünk a több milliónyi adatból, amelyet egy nagyobb méretű nemesítési program generál évről-évre.

Az előadás a martonvásári búzanemesítéssel foglalkozó mezőgazdasági kutatás céljára létrehozott információs rendszert mutatja be. A kidolgozott technológia célja a kutatói közösség nemesítési és szántóföldi kísérletezésekkel kapcsolatos közös ismeretrendszerének egységbe foglalása és az egyes tevékenységek információtechnológiai támogatása. A kidolgozott rendszer rendkívül hatékony a modern búzanemesítési kutatási programok menedzselésének tökéletesítésében, mert lehetővé teszi a kutatási programok méreteinek további növelését (a humán erőforrások létszámának csökkentése mellett is), támogatja a csoportos tevékenységek lebonyolítását, biztosítja a teljes adatkonzisztenciát és végül az eredmények feldolgozásával adatokat szolgáltat a kutatási program teljesítményének jellemzésére.

Az adatkezeléshez és kiértékeléshez korábban használt szövegszerkesztők, táblázatkezelők, adatbázis-kezelők, statisztikai programok helyett egy integrált adatmodellre támaszkodó egységes alkalmazáscsomag (*Breeder*) nyújt segítséget a felhasználóknak tevékenységeik végzéséhez.

A rendelkezésre álló legfontosabbak eszközök lehetővé teszik a vonalkód alapú automatizált adatgyűjtést, változó szerkezetű listák és különféle outputok előállítását, összetett szántóföldi kísérletezési és nemesítési műveletek szervezését, teljes körű pedigre és génbanki nyilvántartás létrehozását és felügyeletét, az alapanyagcsere szervezését és a hozzátartozó címlista adatnyilvántartását, alapvető elemzések készítésére alkalmas statisztikák futtatását.

Az előadásban bemutatott új technológia révén az alkalmazott informatika jelentős mértékben válhatott a tevékenységek szervezése, az adatfeldolgozás és kiértékelés során a hatékonyság és a költségtakarékosság növelésének fontos eszközévé. A rendszer felhasználói köre a búzanemesítési és szántóföldi tevékenységeket végző kutatókra, intézeti mérnökökre, technikusokra és fizikai dolgozókra terjed ki elsősorban.

Kulcsszavak: információs rendszer, agrárinformatika, nemesítés, szoftver, alkalmazás

Bevezetés

Az elmúlt két évtizedben a számítógépek használata a mezőgazdasági kutatással foglalkozók körében gyorsan terjedt és ezzel egyidőben kezdtek kialakulni az adatkezelési szokások is. Az általunk vizsgált nemzetközi gyakorlat és a hazai tapasztalatok is azt mutatják, hogy speciális szoftverek hiányában, a mezőgazdasági kutatásban a kis és nagyobb szervezetek egyaránt kénytelenek voltak „egyéni” megoldásokat kifejleszteni az évek során. Az adatkezelés és kiértékelés a legtöbb esetben a számítógépeken szövegszerkesztők, táblázatkezelők, adatbázis-kezelők, statisztikai programok segítségével történt.

A kutatási programok méretének növekedése, amely általában együtt járt a humán infrastruktúra optimalizálásával, megmutatta, hogy ezek a megoldások nem képesek a kutatási tevékenységek szervezésére, nagy adatszám esetén az adatok feldolgozása nehéz, a nem strukturált adatok közötti összefüggések vizsgálata nem lehetséges, konkrét helyzetekben a döntés az azt alátámasztó

információk nélkül bizonytalan, az elemzések, jelentések elkészítése mindennapos nehézségekbe ütközik.

Szükségessé vált olyan technológia kifejlesztése, amely egységbe integrálja a komplex forrásokból származó különféle adatokat, leír és eloszt tevékenységeket, meghatározza ezek időbeni sorrendjét és végül az eredmények feldolgozásával adatokat szolgáltat a kutatási program teljesítményének jellemzésére.

Általánosságban elmondhatjuk, hogy még napjainkban is igaz az, hogy amíg az alapszoftvereket a szoftverpiacról lehet és célszerű beszerezni, addig a speciális mezőgazdasági kutatások és alkalmazási területek igényeit kielégítő felhasználói szoftverek, és információs rendszerek jelentős részét még mindig egyedileg kell kifejleszteni.

Irodalmi áttekintés

Információs technológia

Az információs technológia (Information Technology - IT) egy gyűjtőfogalom, felöleli azon technológiák minden formáját, amelyek használatosak az információ létrehozására, tárolására, kicserélésére és felhasználására (Soanes et al., 2005). Olyan tevékenységeket jelent, mint alkalmazástervezés, információs rendszerek (Information System - IS) tervezése, menedzselése, felhasználóképzés és támogatás. Lehet túlzottan technikai jellegű, de lehet kifejezetten emberközelit is, célja túlnyomó részben a számítástechnikai ismeretek gyakorlatban történő alkalmazása az élet számos területén. Az IT lehetőségeit kihasználva rendszerszervezési célokat valósíthatunk meg. Leegyszerűsíthetjük bizonyos feladatok elvégzését a szervezetek számára anélkül, hogy a felhasználónak törődnie kellene azzal, hogy mi is történik a háttérben (Senn, 2004).

Ennek eredményeképp jöhettek létre az információs rendszerek (IS), amelyeknek elsődleges célja a hardveren tárolt adatokból szoftverek és egyéb alkalmazói programok segítségével lehetővé tenni a felhasználó számára az információ gyors, pontos kinyerését.

Információs rendszerek általános jellemzői

Az adatok mennyiségének exponenciális növekedésével nagyon sok helyen gondot okoz az ún. „gazdag adat, de szegényes információ” szindróma. Ennek következménye, hogy a költségesen előállított és rögzített adatok nem tudnak hozzájárulni megalapozott döntések előkészítéséhez.

Feltételezve, hogy egy kommunikáló közösségnek van közös, egységesen értelmezett ismeretrendszere, minden új üzenet a közös ismeretekhez viszonyított relatív üzenetnek tekinthető a közösség szervezett kommunikációján belül (Benczúr, 2000). Amennyiben az üzenet nem csupán átfut a közösségen, hanem hozzáadódik, beépül a közös ismeretrendszerbe, gondoskodni kell arról, hogy ezt mindenki (vagy akikre tartozik) megismerhesse, és ne tudjon hozzáférni az, akire nem tartozik. Az információs rendszerek ezt a szerepet töltik be.

Az információs rendszerek az információk tartalma, köre, jellege szerint sokfélék lehetnek, pl. országos, ágazati, vállalati, statisztikai, mezőgazdasági, stb. (Szép, 2004). Az információs rendszerek összegyűjtik, feldolgozzák, tárolják, visszakeresik, továbbítják a különböző szervezetek működésével kapcsolatos adatokat, információt állítanak elő, felméri a működéshez szükséges erőforrásokat, meghatározzák a rendelkezésre álló erőforrások allokálását és segítik a vezetői döntéshozatalt (Hetyei, 2002).

Az információs rendszerek nagy előnye, hogy a tudatos szervezés eredményeként egy adott közösség kommunikációjának egy részét szabályozza, formalizálja. Ezzel az új lehetőségrendszerrel az információs rendszerben résztvevő közösség számára az információcsere nagy része a számítástechnikai és kommunikációs (hálózat) rendszer közbeiktatásával történik.

A modern információs rendszerek, támaszkodva az adat, szoftver, hardver, humán és kommunikációs erőforrásokra, megvalósítják a bevitel, feldolgozás, kimenet, tárolás és felügyeleti

tevékenységeket, amelyeknek eredményeképpen a forrásadatokból a végfelhasználó számára hasznos információk, eredmények jönnek létre.

Adatmodell, adatbázis

Az információs rendszerek passzív eleme az adatstruktúra. Csak jól megtervezett adatsztruktúrára épülő információs rendszer működtethető hatékonyan, mert így biztosítható a tárolt adatok konzisztenciája, épsége és pontossága. A jó tervezéssel hosszú távon időt takaríthatunk meg, mert: (1) az adatstruktúra szerkezete könnyen módosítható és karbantartható lesz, (2) az adatok módosítása könnyen elvégezhető, (3) könnyű kinyerni az adatokat, (4) végfelhasználói alkalmazások fejlesztése egyszerű feladat (Hernandez, 2004).

Az adatbázisokból, az adatbázisokat kezelő programokból (adatbáziskezelő, speciális felhasználói alkalmazások) kialakított adatbázisrendszer létrehozása ill. üzemeltetése egy szervezetnél azt jelenti, hogy az adattömeg egy központi erőforrássá válik.

Az adatbázisok köznapi értelemben valamely rendezett, valamilyen szisztéma szerint tárolt adatokat, az adatoknak kapcsolataikkal együtt való ábrázolását, tárolását értjük (Kovács, 2004). A kapcsolatok a valóságmodell szerves részei, azok az objektumok, amelyek kapcsolatban vannak egymással a valós világban, az adatbázisban is kapcsolatban lesznek, tehát az adatbázisban a kapcsolatok is tükröződnek (Fábián, 2005). Mit érne egy olyan nyilvántartás, amelyben mind a genotípusok, mind a minőségadatok benne lennének, de a rendszer nem tárolná, hogy mely tulajdonságok mely genotípusra jellemzőek.

Az adatbázisok szerkezetének modelljét az adatmodellezés fázisában készítjük el. Az adatmodellezés az adatállományok (fájlok) tervezésének korszerű módszere (Date, 2003). Az adatbázisrendszerek modelljéül szolgáló struktúrák közül ma a legnépszerűbb, a tiszta elméleti alapokon nyugvó relációs adatmodell (Codd, 1970).

Számítógépek a mezőgazdaságban

A mezőgazdasági szoftverek népszerű tematikája a nemesítés és az azzal bizonyos mértékben összefüggő farmgazdálkodás. Az elmúlt két évtizedben nemesítéssel kapcsolatos szoftver fejlesztésekről az *Agronomy Journal* (Agronomy Journal Online, 2004) több mint 40 cikket, a *Crop Science* (American Society of Agronomy) 69 cikket közölt, a *Plant Breeding* (Blackwell Publishing) megjelent 124 ilyen témájú cikk. Ezek zöme növénymodellezéssel illetve géntechnológiával kapcsolatos. Az áttekintett szakfolyóiratokban hagyományos búzanemesítéshez kapcsolódó adatmodellekkel és hozzájuk tartozó alkalmazások fejlesztéséről egy helyütt történik említés az *Euphytica*-ban (Láng et al., 2001).

A témakörben megjelent több száz szoftver közt jórészt olyanokat találunk, amelyek egy-egy kutatási részterület tevékenységeinek problémáit, feladatait segítenek megoldani. Ilyenek pl. a szelekció (Lauer, 1995), minőség (Haley et al., 1999), betegség-ellenállóság (Audsley et al., 1997), fajta információs rendszer (Shroyer et al., 1987), adatgyűjtés (aratósi rendszerek), fajtabejelentés és nemesítói jogok (Albert et al., 1997) témakörében létrehozott megoldások. Ezekon kívül vannak még specializált - egyetlen növényre kidolgozott - nemesítési rendszerek: burgonya (Kozub et al., 2000), zöldségfélék (Bisby, 1989), édesburgonya, (St. Amand et al., 1992), stb.

Az áttekintett szoftverek között kevés olyat találtunk, amely a nemesítő szelekciós döntéseinek támogatását hivatott elősegíteni - valószínűleg azért, mert ez érdeemben csak egy átfogó és integrált adat háttérrendszer támogatásával lehetséges.

Búzanemesítés számítógépesítésének korábbi eredményei Martonvásáron

A nemesítési anyag tudatos kezelésére, az összetett genealógiai információk hasznosítására csak akkor van lehetőség, ha azt egy megfelelő nyilvántartási rendszer segíti. Az áttérés a búza adatok számítógépen való tárolására a nyolcvanas évek közepén kezdődött Martonvásáron. A

nemesítési anyag jelentős kiszélesedése és az újabb igények megjelenése megsokszorozták a korábbi adatmennyiséget, miközben a költséges, hibákkal terhelt kézi feldolgozás egyre nehezebben felelt meg a követelményeknek.

A martonvásári számítógépes fejlesztések korai ún. „adatrögzítő” szakaszában főleg a pedigre adatok tárolásának megkönnyítésére alkalmaztak először számítógépet (COMMODORE-64) a búzakutatásban (programozó Ábrányi A., nem publikált). A két évtizeddel korábbi „alkalmazott számítástechnikának” fontos eredménye volt, hogy megszüntette a pedigrek - korábban hetekig tartó - kézzel történő másolását, utólag pedig a rögzített adatokat át lehetett tölteni az időközben megjelent asztali számítógépekre.

Az asztali számítógépek elterjedése és azok az új lehetőségek, amelyeket kínált, szinte azonnal elindították azt a folyamatot, melynek eredménye a martonvásári számítógépes fejlesztések második szakaszában (1988-1989) megtervezett és létrehozott ún. „nemesítési programcsomag”. Ennek keretében elkészültek a kísérleti és genealógiai adatokat tartalmazó első egyszerű adatbázisok (dBASE III, FoxBase adatbázis-kezelők) és ezzel egyidőben születtek meg az első alkalmazások (Clipper nyelven). Ezek a programmodulok a nemesítési alaptevékenységek egy részének adatkezelését tették lehetővé (szelekció, keresztezés, magérkezés- és magküldés, parcellázás, stb.) és a szántóföldi műveletek kivitelezéséhez is praktikus eszközöket biztosítottak (programozó Kuti Cs., nem publikált).

Ekkor még nemzetközi viszonylatban is csak kevés szoftver létezett a témakörben, ezek jobbra különféle kísérleti elrendezések megválasztásához és statisztikai kiértékelések elvégzéséhez nyújtottak segítséget, kihasználva a PC-k megjelenése által hirtelen megnövekedett számítási kapacitást. Ilyen volt, pl. a Basic-ben íródott ANOFT (Pesak, 1991) vagy az MSTAT-C (1988).

Anyag és módszer

Szoftver

A „Breeder” alkalmazáscsomagot a Microsoft® VB-IDE (Visual Basic Integrated Development Environment) általános célú programfejlesztői környezetének segítségével fejlesztettük. Mintegy 86 MB-nyi területet foglal el a fejlesztő gépén, a felhasználók gépeire ezt nem kell telepíteni.

A „Breeder” által létrehozott (output) és a „Breeder”-be bemenő (input) adatok többsége olyan formátumokban van, amely a Windows® alatt futó Microsoft Office2000® és OfficeXP® csomagok részét képező Microsoft® Excel, Microsoft® Access és Microsoft® Word alkalmazásokban is megjeleníthető, így az optimális működéshez célszerű az Office2000/XP irodai csomagok valamelyikét telepíteni a munkaállomásokra (kb. 150 MB).

Hardver

Munkaállomások

Azok a számítógépek, amelyeket a nemesítési információs rendszer fejlesztői és felhasználói (kutatók, intézeti mérnökök, technikusok) munkájuk során használnak. Tipikusnak mondott konfiguráció: Intel Pentium IV 1,5 GHz-től (vagy kompatibilis), memória: 256 MB RAM, merevlemez: 160 GB, monitor: SVGA 17”, hálózati kártya (10/100 Ethernet LAN), optikai egység (CD/DVD), USB csatlakozó, egér, billentyűzet.

Automata adatgyűjtők

Azokat a számítógépeket, amelyek még működőképesek, de teljesítményük a mindennapi munkavégzéshez már nem elegendő. Tipikusnak mondott konfiguráció: Intel Pentium III 800 MHz-től (vagy kompatibilis), memória: 128 MB RAM, merevlemez: 40 GB, monitor: SVGA 15” (800x600), hálózati kártya (10/100 Ethernet LAN), optikai egység (CD/DVD), soros (RS232) csatlakozó, egér, billentyűzet.

Hálózati eszközök

A munkaállomásokról érkező kérések kiszolgálásának biztosítására egy IBM eSeries 200 típusú központi adatbázisszerver üzemel az alábbi kiépítésben: 1,6 GHz-es processzor, 160 Gbyte-os merevlemez, 1 Gbyte RAM-al felszerelve.

Adatfájlok és tárigény

Az 1984-től kezdődően évente létrehozott adatbázisok együttes mérete kb. 200 MB. Ezek a fájlok nem egyforma méretűek, hanem arányban állnak a nemesítési anyag növekedésével és a gyűjtött adatok típusának és számának évről-évre történő gyarapodásával. A felsoroltakon kívül még tárolunk egyszerűbb szöveges és Excel típusú alapfájlokat, paraméterfájlokat, ezekben kísérleti elrendezések, programbeállítások, eltárolt címkeformátumok vannak.

Vonalkódalapú eszközök

Címketípusok

A vonalkódok nyomtatása a kereskedelmi forgalomban kapható szabványos méretű öntapadós és műanyag címkékre történik. Az öntapadós címkéket hagyományos lézer (HP, CANON) nyomtatókkal nyomtatjuk Avery-Zweckform típusú szabványos (AZ3651, AZ3652) címkeívekre. Az általunk használt thermo-érzékeny műanyagcímkék speciális anyagból készülnek (HDPE), 0,25mm vastagságúak, kereszt alakú részettel, kétféle méretben: 11cm x 2,5cm, ill. 11cm x 5cm.

Vonalkódnymtatók

A műanyag címkék nyomtatása speciális eszközök igénybevételével történik. Az általunk használt thermo-transzfer nyomtatók: TTX300 COBRA, TTX350 OCELOT (Avery Dennison Corporation, 1985) max. 110mm széles címkék előállítására alkalmasak. Műszaki szempontból jelentősen eltérnek a hagyományos nyomtatóktól.

Vonalkódleolvasók

A mérési adatok automatikus azonosítása CCD (Charge Coupled Devices) típusú optikai leolvasókkal történik: BCH5X49, BS-L01

Adatgyűjtési eszközök

- NIR/NIT készülékek: FOSS Infratec® 1241, Perten Inframatic Ö 8611,
- digitális mérlegcsalád (Mettler): PB602-S, Viper SW, Spider SW,
- szemkeménységmérő készülék: Perten SKCS 4100,
- dagasztott tészta tágulásának vizsgálata (Chopin): Alveograph, Alveolink,
- reológiai tulajdonságok mérése (Brabender): Farinograph, Farinograph-E,
- textúra elemző rendszer (Stable Micro Systems): TA.XTPlus,
- kombájnrá szerelhető automata adatgyűjtők: OMNIDATA Polycorder, HARVESTMASTER HM-Fieldbook,

Alkalmazott statisztikai módszerek és kísérleti elrendezések

Figyelembe véve, hogy az egy helyen, egy táblában beállított többismétléses kísérlet adja a fajta teljesítményének az adott táblára vonatkozó legmegbízhatóbb becslését (Matuz, 1987), az egytényezős többismétléses és többtényezős többismétléses kísérletek statisztikai értékelését beépítettük a rendszerünkbe. A véletlen blokkelrendezésű egytényezős és különböző elrendezésű kéttényezős (véletlen blokk, osztott parcellás) kísérletek statisztikai kiértékelését a Sváb által leírt módszer (Sváb, 1981) alapján végezzük és értelmezzük.

A kéttényezős varianciaanalízis esetében a két minőségi tényező kölcsönhatásának megléte vagy hiányának kimutatása fontos lehet számunkra. A kéttényezős variancia tábla értékelését szignifikáns kölcsönhatás esetén a Tóthné által leírt módszer (Tóthné, 2004) alapján végezzük és értelmezzük.

Az öntermékenyülő gabonafajok nemesítésére leggyakrabban alkalmazott nemesítési eljárások (pedigré- és tömegszelekciós módszer; Bedő, 2004) közül a pedigrémódszer sajátosságaihoz optimalizáltuk az adatszerkezetet.

Eredmények és következtetések

A nemesítési és szántóföldi kutatási program méretének és hatékonyságának növeléséhez információs infrastrukturális háttérrel hoztuk létre. Egy áttekinthető rendszert építettünk fel, amely az általánosságra való törekvés mellett, tükrözi a Martonvásáron sok éves tradíció során kialakult kutatási szemléletet. A rendszer alkalmas a nemesítési anyag nyilvántartására, szántóföldi és laboratóriumi kísérletek tervezésére, a kísérleti adatok automatizált gyűjtésére, és a kísérletek gyors kiértékelésére. Az intézet gabona génbankjának nyilvántartása és az alapanyagcsere a nemesítési adatokkal egységes rendszerben történik.

Relációs adatmodell megtervezése és az adatstruktúra létrehozása

A martonvásári búzanemesítési adatstruktúra része az információs rendszernek. Célja (1) a nemesítéssel kapcsolatos összes adat egységes, integrált nyilvántartása, (2) döntéstámogató információk, jelentések csoportosításának és teljes körű megjelenítésének lehetővé tétele, (3) a nemesítési adatok kapcsolatának biztosítása a genealógiai, génbanki és alapanyagcsere nyilvántartásokkal.

Kiindulásként elemeztük a rendelkezésre álló, meglévő adatbázisainkat. Ezek kisebb részben „papír”, túlnyomó részben „hagyaték” alapú adatbázisok voltak. A „hagyaték” adatbázisok esetében olyan adatbázisokról volt szó, amelyek már régóta használatban voltak a búzanemesítésben, tartalmukat különféle PC-alapú tárolási technikák segítségével rögzítették: táblázatkezelők, szövegszerkesztők, adatbázis-kezelők. Ezek közül mind méretüket, mind tartalmukat illetően a legfontosabbak voltak az 1984 óta meglévő egyszerű nemesítési adatállományok. A korábbi adatszerkezési módszereknek megfelelően ezek az évente megújuló „mindentudó” (flat) fájlok tartalmazták tulajdonképpen az illető évre vonatkozó összes fontos adatot, köztük a különféle kísérletek alapadatait, genealógiai adatokat, megfigyelési adatokat, beltartalmi és egyéb mérési eredményeket.

Adatainkat témakörök szerint csoportosítottuk, így jöttek létre az adatmodellen belül az egyes adatbázisok: az évről-évre megújuló nemesítési adatbázisok, a genealógiai (pedigré) adatbázis, az alapanyagcsere adatbázis, génbanki adatbázis. Az egyes adatbázisokban helyet foglaló táblák számát az adott témakör által determinált témák száma határozza meg, így a nemesítési adatbázisban 12 tábla van. Külön táblákban helyezkednek el a kísérletek, a szántóföldi megfigyelési adatok, a minőségvizsgálatok eredményei, súlynyilvántartás, stb.

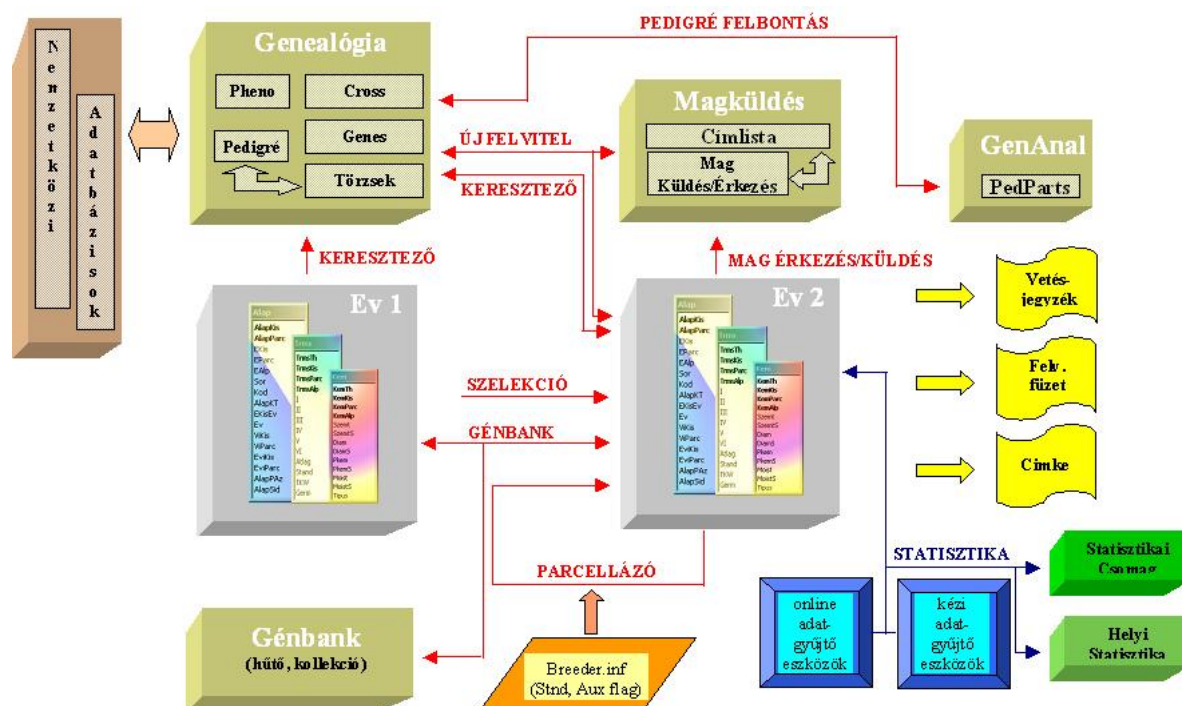
A genealógiai adatbázisban tároljuk genotípusok pedigréit, fajta- és törzsneveit, faj- és életforma adatait valamint a fajtabejelentésre vonatkozó adatokat. Ugyancsak itt kaptak helyet bizonyos fenotípusos adatok, keresztezési listák és a molekuláris genetikai adatok külön-külön táblákban.

Az alapanyagcserével kapcsolatos témakör három nagy témából tevődik össze: Magérkezés, Magküldés és Címlista. A Génbanki nyilvántartás egyetlen témakörből áll: a génbankban tárolt anyagok és minták adatait tartalmazó „KOLL” nevű táblából.

Tevékenységszervező programmodulok

Az adatstruktúra az információs rendszer passzív eleme, ahhoz hogy ezek az adatok felhasználhatóvá váljanak algoritmusok, alkalmazások kellenek, mellyel ezek az adatok, információk életre kelthetők. A nemesítés során keletkező adatok kezelésére, új adatstruktúrák létrehozására, a nemesítés körébe tartozó tevékenységek szervezésére, különféle döntések előkészítésére, statisztikai elemzések végzésére alkalmazáscsomagot (*Breeder*) fejlesztettünk ki.

A Breeder felhasználói felülete közel ötven - az egyes tevékenységek technológiai leírásán alapuló - alkalmazást foglal egységes keretbe. Tartalmaz menüket, almenüket és egy gyorsindító eszköztárat a leggyakrabban használt programmodulok meghívására.



1. ábra. Az információs rendszer összefoglaló blokk diagramja

Az 1. ábra blokkdiagramján láthatók az információs rendszer egyes elemei: az adatbázisok, a fontosabb outputok, az automata adatgyűjtő modul, a statisztikai modul és a nyilakkal jelölt legfontosabb alkalmazások, amelyek összekapcsolják az egyes részeket.

A leggyakrabban futtatott modulok között első helyen kell említenünk a kísérletek tervezéséhez segítséget nyújtó alkalmazásokat: a keresztező, szelektáló és parcellázó modulokat.

Keresztezéskor a program automatikusan hozza létre az új pedigrét (Purdy et al., 1968), a szelekció történetét rögzítő „Kód” információt és keresztezési listákba menti az egyes keresztezési programokkal kapcsolatos fontos adatokat. Ezekből később elő lehet keresni az anyai és a beporzó összesítőket, tudni lehet hogy a szülőket mely kísérletekből kell elővenni és azt is, hogy az illető szülő összesen hányszor szerepel a keresztezési programban.

A szelekciós modul rögzíti a genotípus származását (előző évi: kísérlet, parcella), az elvetett sorok számát, automatikusan ír: kódot (szelekció történet), új törzsazonosítószámot (ha szükséges) és kapcsolatot tart a fajtafenntartással, fagy- és rezisztencia kísérletekkel. Automatikusan nyilvántartja az egyes genotípusok vetőmagjaiból rendelkezésre álló mennyiségeket és lehetőség van - az egyenkénti kijelölés és átszelektálás helyett - akár több száz genotípust is automatikusan átszelektálni, előre beállított bonitálási értékek és egyéb feltételek megadásával.

A parcellázó modul nemcsak a több szempont alapján történő végső parcellasorrend kialakításának, hanem igény szerint bizonyos kísérletek kontrollokkal való kiegészítésének, egyéb esetekben előre megadott méretű kísérletek automatikus előállításának hatékony eszköze is. Ez attól függ, hogy a rendelkezésre álló (négyféle) parcellázási módok közül a felhasználó melyiket választja. A végleges sorrend kialakításakor a nemesítőnek lehetősége van elkülöníteni a különböző kalásztípushoz tartozó törzseket, sorba rakni őket (pl. az előző évi kalászolási idő szerint), azonos kalászolási idejű genotípusokon belül a testvér törzseket csoportosítani, stb.

A gyorsindító eszköztárról kezdeményezhető a génbanki nyilvántartást végző modul, amelynek segítségével megvalósítjuk a génbanki és a nemesítési adatbázisok között zajló adatcserét (tárolás,

kitermesztés). A génbanki modul a kiválasztott kísérlet genotípusait mindig szelektíven ajánlja fel az éppen aktuális műveletek végzéséhez. Ha egy adott kísérletből (kísérletcsoportból) új tételeket szeretnénk felvinni a génbankba, akkor csak azokat a genotípusokat ajánlja fel, amelyek még nem szerepelnek a génbankban. Ellenkező esetben, amikor arra vagyunk kíváncsiak, hogy egy adott kísérlet genotípusai közül melyek azok, amelyekkel frissíteni tudjuk a génbank egy adott célkísérletét, akkor a kísérletből csak azokat a genotípusokat látjuk, amelyek már jelen vannak ott.

Szintén a gyorsindító eszköztárról indítjuk a nemesítési és génbanki alapanyagok bel- és külföldi elosztását és az onnan érkező tételek felügyeletét végző modult. Ennek segítségével tartjuk nyilván úgy a beérkező, mint az intézetből kiküldött magtételekhez kapcsolódó összes információt.

A gyakran futtatott modulok közé tartozik a felvételezési adatok kézi bevitelét lehetővé tevő modul, a külső (Excel) adatfájlok tartalmát az információs rendszer adatstruktúrájába illesztő modul, a genealógiai adatok konzisztenciáját ellenőrző modul, a vetésterveket és felvételezési füzeteket előállító modul, a vonalkódok tervezését és nyomtatását végző modul, a különböző laboratóriumi minőségvizsgáló műszerekről és automata adatgyűjtővel ellátott kombájnokról származó adatokat automatikusan gyűjtő modul.

Pedigré nyilvántartás

A nemesítési és szántóföldi adatok nyilvántartásának sarokpontja az egységes pedigré model. Az egységes pedigré model központi gondolata a genotípusok egyedi azonosítók hozzárendelésével történő azonosítása, a homonímák és szinonímák kezelése, amelyek természetesen módon adódnak a genotípusok fejlődése és használata során.

A jelenleg több mint 110.000 különböző genotípus elnevezést tartalmazó nyilvántartásunk számára létrehoztunk egy különálló adatstruktúrát. Minden pedigré bejegyzéshez két azonosítót rendelnek automatikusan a pedigré adatokat kezelő programmodulok. Az első (PAz) az eltérő kombinációkat azonosítja, a második (SID) az azonos keresztezési kombinációból származó testvértörzseket különbözteti meg egymástól.

Minden genotípus elnevezés csak egyszer, egy központi nyilvántartásban szerepel, és a genotípusra vonatkozó genealógiai információk innen csatolhatók a különböző évek, vagy kísérletek nyilvántartásához. Azoknál a genotípusoknál, amelyeknél több elnevezés is van forgalomban (pedigré, törzsnév, fajtanév, stb.) az adatbázist kezelő programmodulok eldöntik, hogy ezek közül melyik a legaktuálisabb és azt mutatják a felhasználónak.

Adatgyűjtés

Amikor adatgyűjtésről beszélünk elsősorban az automatizált, vagy félig automatizált funkciók megvalósítására gondolunk. Az adatgyűjtés automatizálásához szükség van olyan korszerű technikák alkalmazására, amelyek a mikroprocesszorral felszerelt és számítógéphez csatlakoztatható digitális vezérlésű mérési- és analitikai eszközök, széleskörű elterjedésével elérhetővé váltak. Az adatgyűjtési funkciók automatizálásával az adatgyűjtés és az adatfeldolgozás üteme (és minősége) azonos szintre emelhető, ez nagyon fontos olyan esetekben, ahol a megtakarított időt tekintjük meghatározó tényezőnek (pl. aratás és vetés közötti rövid időköz optimális kihasználásakor).

Az automatizált adatgyűjtéshez alapvetően két dolog szükséges: kommunikáció a megfelelő műszerrel és az automata adatazonosítás.

A kommunikációhoz műszerspecifikus szoftveres interfész szükséges, amelyet minden műszerhez, vagy műszer családhoz külön létre kell hozni. A befektetett munka megtérül, mert az adatok hibátlan, gyors gyűjtése a műszer hátralévő élettartamára megoldott.

Rendszerünkben az automata adatazonosítás alapja a vonalkód, ezért beépített vonalkódtervező és nyomtató modult hoztunk létre. Ez több adatmezőt is össze tud vonni egyetlen vonalkódba, és ami nagyon fontos, az adatok közvetlenül az adatbázisból nyomtathatók öntapadós vagy műanyag alapú címkékre.

Az általunk alkalmazott vonalkódokkal ellátott címkéknek kettős szerepe van: egyrészt a humán olvasásra alkalmas felirat rendelkezésre áll felvételezéskor/vetőmag előkészítésekor, másrészt az alatta elhelyezkedő vonalkódokat a mérési eredmények adatbázisokba történő elhelyezéséhez használjuk azonosításhoz.

Statisztikai modul

Bizonyos kutatási programok számára csak egy szűken kimért időkeret áll rendelkezésre a kutatások szakmai eredményeinek kiértékelésére, különösen igaz ez a nemesítési programok esetében, amelyeknek az aratási szezon végén általában napok állnak rendelkezésre a jelentések elkészítéséhez és hetek alatt kell előkészíteni a következő ciklus tenyészkeretekben elvetésre kerülő anyagait.

Az egy vagy több termőhelyen elvetett azonos vagy különböző randomizációjú kísérletek alapvető statisztikai elemzéseinek elvégzéséhez nincsen szükség külön statisztikai csomagok használatára, sem az adatok előkészítésére ezekhez az elemzésekhez. A rendszere beépített alapvető statisztikák - leíró statisztika, véletlenblokk varianciaanalízis (egytényezős, kéttényezős), osztott parcellás (split-plot) elrendezés, lineáris regresszióanalízis (egyszeres, többszörös), korrelációs mátrix számítása, igen egyszerűen futtathatók közvetlenül a központi adatbázisból.

Következtetések és javaslatok

A búzanemesítéssel foglalkozó kutatás versenyképességének fontos feltétele a kutatási programok méretének növelhetősége, az egyre nagyobb számban keletkező adatok hatékony kezelése, de mindenekelőtt a hasznos, döntéstámogatáshoz nélkülözhetetlen információk szintetizálása a több milliónyi adatból, amelyet egy nagyobb méretű kutatási program generál évről-évre.

Információs rendszerünk létrehozásakor abból indultunk ki, hogy egy nagyméretű kutatási programban csapatmunka folyik. Egyidejűleg többen dolgoznak a feladatok megvalósításán, azt többször is megszakítva más időszereű részfeladatok kényszerű beütemezése miatt. A rendszertervező számára ez azt jelenti, hogy minél nagyobb szabadságot kell biztosítani a kutatóknak és a technikai személyzetnek ahhoz, hogy közös tevékenységeiket különböző időpontokban tudják végezni. A rendszer feladata fenntartani a mindenkori változtatások és a megkezdett műveletek folytatásának lehetőségét úgy, hogy előreláthatólag ne kelljen semmit előről kezdeni. Ebből adódik egy lényeges különbség, ami a rendszerünk és a vizsgált szoftveres megoldások között van. Utóbbiaknál már a tervezés legelején be kell állítani a kísérletek több olyan jellemzőjét, amelyeknek a rögzítése egy nagyméretű kutatási programban a genotípusok összeválogatásának megkezdésekor még nem áll rendelkezésre vagy jelentős további erőfeszítéseket igényelne.

Az adatszerkezet rugalmassága és az egyes tevékenységek megfelelő szervezése biztosítja számunkra azt, hogy egymásra épülő kutatási tevékenységek elkezdhetők legyenek akkor is, ha az alapul szolgáló műveletek még nincsenek végleges formában kidolgozva. Így pl. nem kell megvárni, amíg egy korai generációs kísérlet több ezer genotípusa véglegesen összeáll, hanem a válogatással párhuzamosan el lehet indítani a már átszelektált genotípusok számára a vonalkódok előállítását és olyan részterveket lehet kinyomtatni, amelyekkel el lehet kezdeni a vetőmag tételek előkészítését is.

A genealógiai adatok felépítése és kezelése úgy a rendelkezésre álló adatok számát, mint a kezelési módszerek bonyolultságát tekintve, fontos szerepet kap a rendszerünkben. Sajátos tulajdonsága a *Breeder*-nek, hogy a nagyméretű nemesítési programok megvalósításához szükséges alapanyag nyilvántartást szoros kapcsolatban kezeli az alapanyagcsere és a génbanki adatbázisokkal.

A *Breeder* statisztikai moduljának célja lehetőséget biztosítani a nemesítési és szántóföldi kísérletek gyors kiértékeléséhez, oly módon, hogy az adatbázisban tárolt mérési és megfigyelési

adatokon közvetlenül és könnyen futtathatók legyenek a statisztikák. Az egy vagy több termőhelyen elvetett azonos vagy különböző randomizációjú kísérletek alapvető statisztikai elemzéseinek elvégzéséhez nincsen szükség külön statisztikai csomagok használatára, sem adatok előkészítésére az elemzésekhez.

Egyedi megoldások a rendszerben: a vonalkód technika beillesztése és az erre alapozott széleskörűen alkalmazott automata adatgyűjtés, a rendszer informatív jellegét biztosító többtáblás, többéves adatlekérdező funkciók, rögzített és változó formátumú és tartalmú vetéselrendezések és felvételező füzetek nyomtatása.

Irodalomjegyzék

- Agronomy Journal Online, (2004): Archive of All Online Issues: 1 Jan. 1998 - 28 Mar. 2005.
<http://agron.scijournals.org/contents-by-date.0.shtml>.
- Albert J., Pfeiffer R. (1997): Data management and analysis with a planning, information and analysis system for field trials (PIAF). *Zeitschrift-für-Agrarinformatik*. 5: 2, 39-43.
- Audsley E, Bailey B. J., Beulah S. A., Maddaford P. J., Parsons D. J., White R. P., Stafford J. V. (1997): Decision Support Systems for arable crops: increasing precision in determining inputs for crop production. *Precision agriculture '97. Volume II. Technology, IT and management. Papers presented at the First European Conference on Precision Agriculture, Warwick University, UK, 7-10 September 1997.* 843-850.
- Avery Dennison Corporation (1985): The professional printers
http://www.machines.averydennison.com/printersystems_gb.nsf/wview/L2L1?OpenDocument.
- Bedő Z., Marton Cs. (2004): Növénynevelési módszerek. 71-100 pp. In Bedő Z. (Szerk): *A vetőmag születése. A vetőmagtermesztés elmélete és gyakorlata.* Agroiinform Kiadó, Budapest, 537 pp.
- Benczúr A. (2000): Informatika – oktatás – informatikaoktatás. *Természet világa. II. különszám.* 86 pp. <http://www.kfki.hu/~cheminfo/TermVil/tv2000/tv0012/00tartp.html>.
- Bisby F. A. (1989): Databases, information systems, and legume research. *Monographs in Systematic Botany from the Missouri Botanical Garden*, 29: 811-825.
- Codd E. F. (1970): A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM* 13 (6), pp. 377-387.
- Date, C. J. (2003): *An introduction to database systems.* Addison-Wesley Publ. Co., Reading, MA. 983 pp.
- Fábián, Z. (2005): *Adatbáziskezelés és Adatbázis szervezés, Jegyzet.* Szily Kálmán Műszaki Középiskola. 56 pp.
- Haley S. D., May R. D., Seabourn B. W., Chung O. K. (1999): Relational database system for summarization and interpretation of hard winter wheat regional quality data. *Crop Science*, 39: 2, 309-315.
- Hernandez J. M. (2004): *Adatbázis-tervezés.* Kiskapu Kiadó, Budapest. 427 pp.
- Hetyei J. (2002): *Pénzüntézetek és állami Intézmények Információs Rendszerei Magyarországon.* Computerbooks, Budapest 388 pp.
- Kovács, L., (2004): *Az adatbázisok tervezésének és kezelésének módszertana.* Computerbooks, Budapest, 482 pp.
- Kozub J. G., Lynch D. R., Kozub G. C., Kawchuk L. M., Fujimoto D. K. (2000): A relational database system for potato breeding programs. *American Journal of Potato Research*, 77: 2, 95-101.
- Lauer J. G. (1995): SELECT!: crop variety selection software for microcomputers. *Journal of Production Agriculture*, 8: 3, 433-437.
- Láng L., Kuti C., Bedő Z. (2001): Computerised data management system for cereal breeding *Euphytica* 119: 1-2, 235-240.

- Matuz J. (1987): Az üzemi kísérletezés jelentősége és problémái. In: Barabás Z. (Szerk.): *A búzatermesztés kézikönyve*, Mezőgazdasági Kiadó, Budapest, 223-236.
- Pesek J. (1991): New procedures in the rationalization of breeding and testing varieties of agricultural crops. *Sbornik-UVTIZ,-Genetika-a-Slechtenti*, 27: 2-3, Priloha, I-V.
- Purdy H. L., Loegering W. Q., Konczak C. F., Peterson C. J. Allan R. E., (1968): A proposed standard method for illustrating pedigrees of small grain varieties. *Crop Science*, 8: 405-406.
- Senn, J. (2004): *Information Technology: Principles, Practices, and Opportunities*. Pearson Prentice Hall. 512 pp.
- Shroyer J. P., Young S. C., Cox T. S. (1987): WHEAT-WIZ: a computer-based cultivar selection tool. *Applied-Agricultural-Research*, 2: 4, 242-247.
- Soanes, C., Hawker, S., (2005): *Compact Oxford English Dictionary of Current English*, Third Edition. 1378 pp.
- St. Amand P. C., Bonte D. R. (1992): A microcomputer database program for sweetpotato breeding and trials. *HortScience*, 27: 3, 268.
- Sváb J. (1981): *Biometriai módszerek a kutatásban*, Mezőgazdasági Kiadó, Budapest. 557 pp.
- Szép K. (2004): A statisztikai információs rendszer és működése. In: Szűcs István (Szerk.) *Alkalmazott statisztika*. Agroinform Kiadó, Budapest, 211-274.
- Tóthné L. K. (2004): Statisztikai hipotézisvizsgálatok. In: Szűcs István (Szerk.) *Alkalmazott statisztika*. Agroinform Kiadó, Budapest, 30-47.