

VALÓS IDŐBEN VÁLASZT ADÓ  
EGÉSZSÉGÜGYI PROFIL  
, MINT TÖBBDIMENZIÓS MEGSZORÍTÁS MÁTRIX,  
ALAPJÁN ÉLELMISZERT SZŰRŐ DOMAIN  
SPECIFIKUS ALGORITMUS

**Kusper Gábor (EKF), Márien Szabolcs (Wit-Sys Zrt.),  
Kovács Emőd (EKF), Kovács László (ME)**

Előadja:

Kusper Gábor

[gkusper@aries.ektf.hu](mailto:gkusper@aries.ektf.hu)

Networkshop 2011, Kaposvár

# Tartalom



- Az EgerFood projekt
- Az eFilter projekt
- Egészségügyi megszorítások
- Megszorítás -> Bitmap Index
- Az IMEE algoritmus
- Adatforrásunk bemutatása
- Teszt eredmények
- Összefoglaló



# EGERFOOD

Egri Regionális Tudásközpont



Keresés > Burgonyás pogácsa

navigációs sáv

Cég: Minta Cég

Név: Burgonyás pogácsa

Szavatossági idő: 2007.10.08.

## Munkalap

Dátum: 2007.10.01.

Gyártósor: Gyártósor

Termelt mennyiség: 160

Termék jelölés: Burgonyás pogácsa

Dagasztási idő: 00:16 ⚠

### Segédanyagok

Anyag: Tartósítószer

Mennyiség: 481

Mennyiség egység: gramm

Anyag: Tojásfehérje por

Mennyiség: 2,53

Mennyiség egység: liter

### Alapanyag

Anyag: (9001) Burgonya

Mennyiség: 36

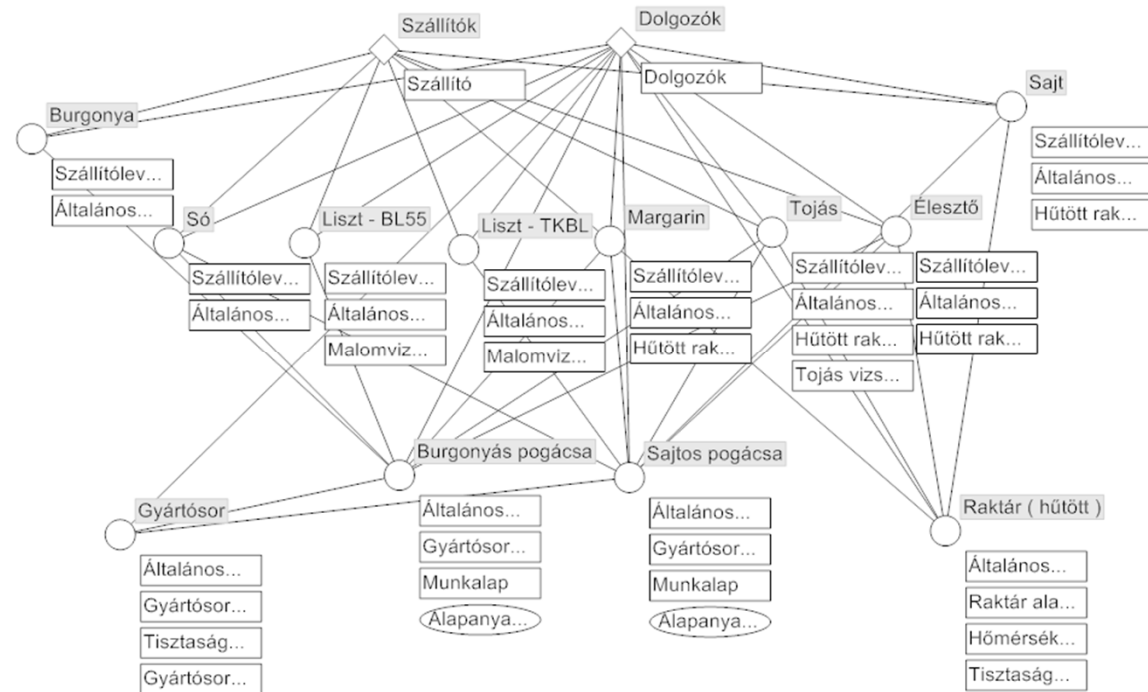
Anyag: (8410) Margarin

Mennyiség: 2

# Az EgerFood információs rendszere



## Munkafolyamat gráf



Az egyedülálló képességek kulcsa a munkafolyamat-gráf. A gráf segítségével minden cég egyedi módon modellezheti a gyártási folyamatait. Ez a modell vezérli a kliens program és az adatbázis működését. A modell szinte végtelen lehetőségeket nyújt és nem mellékesen összetett képet ad a cég működéséről is. Megtervezéséhez ezért a cég képviselőjének és a beüzemelés végző szakemberek közös munkájára van szükség.

# EgerFood cikkek:



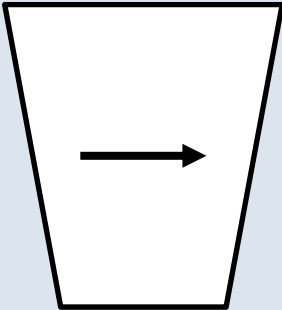
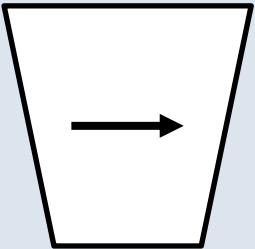
- T. Radványi, G. Kuser: **Requirement Analysis and a Database Model for the Project EgerFood Food Safety Knowledge Center**, ICAI-2007, p. 15-23, 2007.
- K. Liptai, G. Kuser, T. Radványi: **Cryptographycal protocols in the Egerfood Information System**, Annales Mathematicae et Informaticae 34., p. 61-70, 2007.
- K. Gábor, R. Tibor: **Az EgerFood élelmiszerbiztonsági nyomkövető rendszer – Hogyan modellezzük a cégek munkafolyamatait**, Networkshop 2008, Dunaújváros, 8 oldal, 2008.

# Az eFilter projekt



- KMOP-1.1.1-09/1-2009-0053 számú pályázat
- Egészségügyi profil alapján szűrt fogyasztói adatbázisokból nyert információkat kezelő rendszer - eFilter
- WIT-SYS Consulting ZRt.
- Eszterházy Károly Főiskola

# Az eFilter projekt

Bő élelmiszer lista	Egészségügyi profil	Szűk élelmiszer lista
<ul style="list-style-type: none"> <li>•Menük listája</li> <li>•Boltban kapható élelmiszerek listája</li> <li>•Étel rendelésnél étlap</li> <li>•Boltok listája kapható élelmiszerekkel</li> </ul>	 <p>SZŰRÉS</p>	<ul style="list-style-type: none"> <li>•Fogyasztható menük listája</li> <li>•Boltban kapható fogyasztható élelmiszerek listája</li> <li>•Fogyasztható ételek listája</li> <li>•Fogyasztható terméket áruló boltok listája</li> </ul>
Kérdés	Egészségügyi profil	Válasz
<ul style="list-style-type: none"> <li>•Mevásárolt élelmiszer fogyasztható-e?</li> <li>•Vásárlásnál megerősítés, hogy az adott termék fogyasztható-e?</li> </ul>	 <p>ELLENŐRZÉS</p>	<ul style="list-style-type: none"> <li>•Igen / Nem + Megvásárolt élelmiszerről bő információ</li> <li>•Igen/Nem + Megvásárolt élelmiszerről bő információ</li> </ul>

# Egészségügyi profil



- Ételallergiák
- Ételérzékenységek
- Diéták
- Egyéb étkezésnél figyelembe veendő betegségek,  
pl.: cukorbetegség



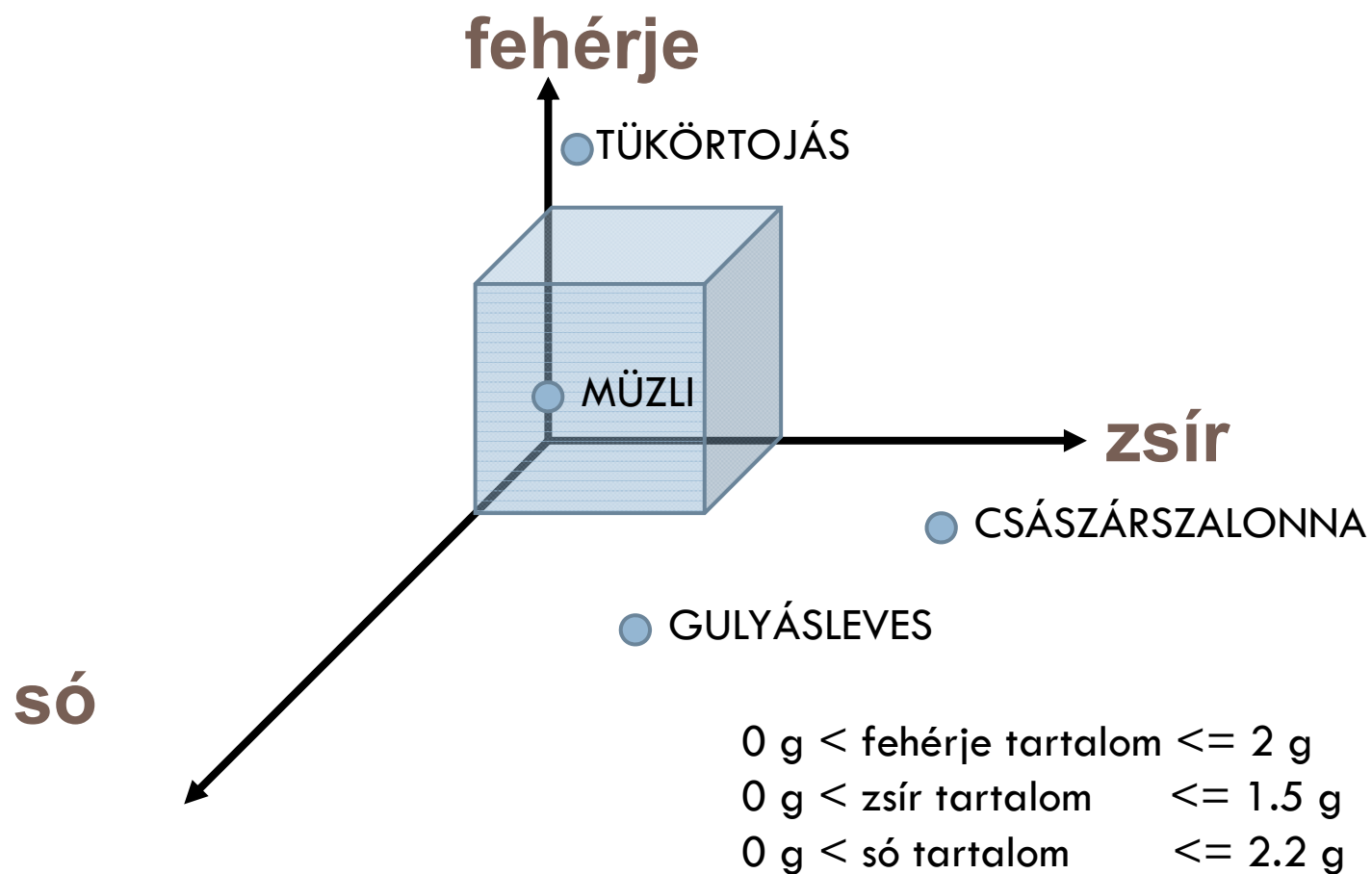
# Megszorítások

- eFilter szabályok típusai:
  - 0 – Tiltás
  - 1 – Nem javasolt
  - 2 – Erősen javasolt
  - 3 – Javasolt
- Étel 100 grammjára vonatkozik.
- Példa:
  - Tiltások: dió > 0g.
  - Nem javasolt: energiatartalom > 500 kcal, zsír > 20g.

# Megszorítások

- Ugyanarra az összetevőre vonatkozó megszorítások összevonhatóak.
- Tiltások:
  - zsír > 30g, zsír > 20g
  - fehérje > 15g, fehérje > 80g
- Tiltások összevonva:
  - zsír > 20g
  - fehérje > 15g

# Többdimenziós megszorítás mátrix



# Megszorítás -> Bitmap Index

- Megszorítás általános alakja:
  - ▣  $N \text{ egység} < \text{összetevő tartalom} \leq M \text{ egység}$ .
- Ötlet: A megszorítás ellenőrzése gyorsítható, ha az összetevőre indexet helyezünk.
- Mivel nagy adatmennyiség esetén sokkal gyorsabb a BITMAP index, ezért használjuk azt.
- Erre akkor van lehetőség, ha az indexelt oszlop a lehetséges értékkészletéből csak néhány értéket használ.

# Megszorítás $\rightarrow$ Bitmap Index

- Megszorítás általános alakja:
  - ▣  $N$  egység  $<$  összetevő tartalom  $\leq M$  egység.
- Határértékek diszkrét értéké tétele:
  - ▣  $f : R \rightarrow B$ , ahol  $|B| \ll |R|$
  - ▣ Jelen előadásban az  $f$  függvény a
    - kerekít( $\log_2(x)$ )
  - ▣ Tehát az indexelt értékek 2 hatványai.

# Megszorítás -> Bitmap Index

- 0. index: Az élelmiszer indexelt, ha az összetevő tartalma nagyobb, mint 0 egység.
- 1. index: Az élelmiszer indexelt, ha az összetevő tartalma nagyobb, mint 1 egység.
- 2. index: Az élelmiszer indexelt, ha az összetevő tartalma nagyobb, mint 2 egység.
- ...
- n. index: Az élelmiszer indexelt, ha az összetevő tartalma nagyobb, mint  $2^{(n-1)}$  egység.

# Megszorítás -> Bitmap Index

- 0. index: Az élelmiszer indexelt, ha az összetevő tartalma nagyobb, mint 0 egység.
- n. index: Az élelmiszer indexelt, ha az összetevő tartalma nagyobb, mint  $2^{(n-1)}$  egység.
- Nézzük néhány szám bitmap indexét:
  - 0 bitmap indexe: ( $>,0,0,0,0$ );
  - 1 bitmap indexe: ( $>,1,0,0,0$ );
  - 2 bitmap indexe: ( $>,1,1,0,0$ );
  - 3 bitmap indexe: ( $>,1,1,1,0$ );
  - 4 bitmap indexe: ( $>,1,1,1,0$ );
  - 5 bitmap indexe: ( $>,1,1,1,1$ );
  - 6 bitmap indexe: ( $>,1,1,1,1$ );

# Megszorítás -> Bitmap Index

- Hogyan tudjuk kifejezni a bitmap index segítségével a tiltást, hogy nem szabad diót ennem?
- Tiltás: dió  $> 0g$ .
- Emlékezzünk vissza az előző dióra:
  - 0. index: Az élelmiszer indexelt, ha az összetevő tartalma nagyobb, mint 0 egység.
  - 0 bitmap indexe:  $(>,0,0,0,0)$ ;
  - 1 bitmap indexe:  $(>,1,0,0,0)$ ;
  - 2 bitmap indexe:  $(>,1,1,0,0)$ ;
  - ...
- Azok a ételek nem ehetők, ahol a 0. index 1-es.
- Azaz dió  $> 0g == (>,1,x,x,x)$  .



# Bitmap Index Maszkok

- 0 bitmap indexe:  
(>,0,0,0,0);
  - (>,0,x,x,x) – összetevő tartalom = 0 egység.
- 1 bitmap indexe:  
(>,1,0,0,0);
  - (>,x,0,x,x) – összetevő tartalom ≤ 1 egység.
- 2 bitmap indexe:  
(>,1,1,0,0);
  - (>,x,x,0,x) – összetevő tartalom ≤ 2 egység.
  - (>,x,x,x,0) – összetevő tartalom ≤ 4 egység.
- 3 bitmap indexe:  
(>,1,1,1,0);
  - (>,1,x,x,x) – összetevő tartalom > 0 egység.
  - (>,x,1,x,x) – összetevő tartalom > 1 egység.
- 4 bitmap indexe:  
(>,1,1,1,0);
  - (>,x,x,1,x) – összetevő tartalom > 2 egység.
  - (>,x,x,x,1) – összetevő tartalom > 4 egység.
- 5 bitmap indexe:  
(>,1,1,1,1);
  - (>,1,0,x,x) – 0 egység < összetevő tartalom ≤ 1 egység.
  - (>,1,x,0,x) – 0 egység < összetevő tartalom ≤ 2 egység.
- 6 bitmap indexe:  
(>,1,1,1,1);
  - (>,1,x,x,0) – 0 egység < összetevő tartalom ≤ 4 egység.
  - (>,x,1,0,x) – 1 egység < összetevő tartalom ≤ 2 egység.
  - (>,x,1,x,0) – 1 egység < összetevő tartalom ≤ 4 egység.
  - (>,x,x,1,0) – 2 egység < összetevő tartalom ≤ 4 egység.

# Kérdés!

- Mi a fehérje bitmap indexe annak a az élelmiszernek, ami 2g fehérjét tartalmaz?
- Tehát mi a 2-nek a bitmap indexe?
  
- Mit kódol a ( $>,1,x,x,x$ ) bitmap index maszk?

# LNTM és LKTM

- LNTM: LegNagyobb Tartalmazott Maszk.
- LKTM: LegKisebb Tartalmazó Maszk.

Megszorítás	LNTM	LKTM
$0,5 \text{ g} < \text{fehérje} \leq 3,5 \text{ g}$	$1 \text{ g} < \text{fehérje} \leq 2 \text{ g}$ (>,x,1,0,x)	$0 \text{ g} < \text{fehérje} \leq 4 \text{ g}$ (>,1,x,x,0)
$1,5 \text{ g} < \text{fehérje} \leq 3,5 \text{ g}$	$2 \text{ g} < \text{fehérje} \leq 2 \text{ g}$ Nem létezik LNTM	$1 \text{ g} < \text{fehérje} \leq 4 \text{ g}$ (>,x,1,x,0)
$N \text{ g} < \text{fehérje} \leq M \text{ g}$	N-nél nagyobb következő indexelt érték < fehérje <= M-nél kisebb indexelt érték	N-nél nagyobb következő indexelt érték < fehérje <= M-nél kisebb indexelt érték
$N \text{ g} < \text{fehérje} \leq M \text{ g}$	$2^{\lceil \log_2(N) \rceil} < \text{fehérje} \leq$	$2^{\lceil \log_2(N) \rceil} < \text{fehérje} \leq$

# Az IMEE algoritmus

- Vonjuk össze az egy összetevőre vonatkozó megszorításokat egy következő alakúra:
  - $N \text{ egység} < \text{összetevő tartalom} \leq M \text{ egység}$
- Minden összevont megszorításhoz határozzuk meg a LNTM.
- Ezen maszkok alapján szűkítjük az adatbázist.

# Az IMEE algoritmus

- Input: Egészségügyi profil, Élelmiszerek adatbázisa minden összetevő kettő hatványára indexelve.
- Output: Tiltott élelmiszerek listája.
- 1. lépés: Az egészségügyi profil tiltó megszorításait összetevőnként összevonom egy „ $N$  egység  $<$  összetevő tartalom  $\leq M$  egység” alakú megszorítássá.
- 2. lépés: Minden összetevőre kiszámolom  $L$  és  $H$  értéket az összevont feltétel alapján:
  - Ha  $N = 0$ , akkor  $L = 0$ , egyébként  $L = \text{padlás}(\log_2(N))+1$ . Ha  $N$  nem létezik, akkor  $L$  sem létezik.
  - Ha  $M = 0$ , akkor  $H = 0$ , egyébként  $H = \text{padló}(\log_2(M))+1$ . Ha  $M$  nem létezik, akkor  $H$  sem létezik.
- 3. lépés: Szűkítem az adatbázist a  $(>,x,\dots,x,1,x,\dots,x,0,x,\dots,x)$  a bitmap index maszkkal, ahol
  - a 1 az  $L$ -dik indexen áll, illetve nincs 1 a maszkban, ha  $L$  nem létezik.
  - az 0 a  $H$ -dik indexen áll, illetve nincs 0 a maszkban, ha  $H$  nem létezik.
- 4. lépés: Ha  $L = H$ , akkor visszaadjuk az üres élelmiszer listát.
- 5. lépés: A szűkítések után megmaradt élelmiszereket visszaadom.

# Adatforrás, 7637 rekord

□ United States Department Of Agriculture:

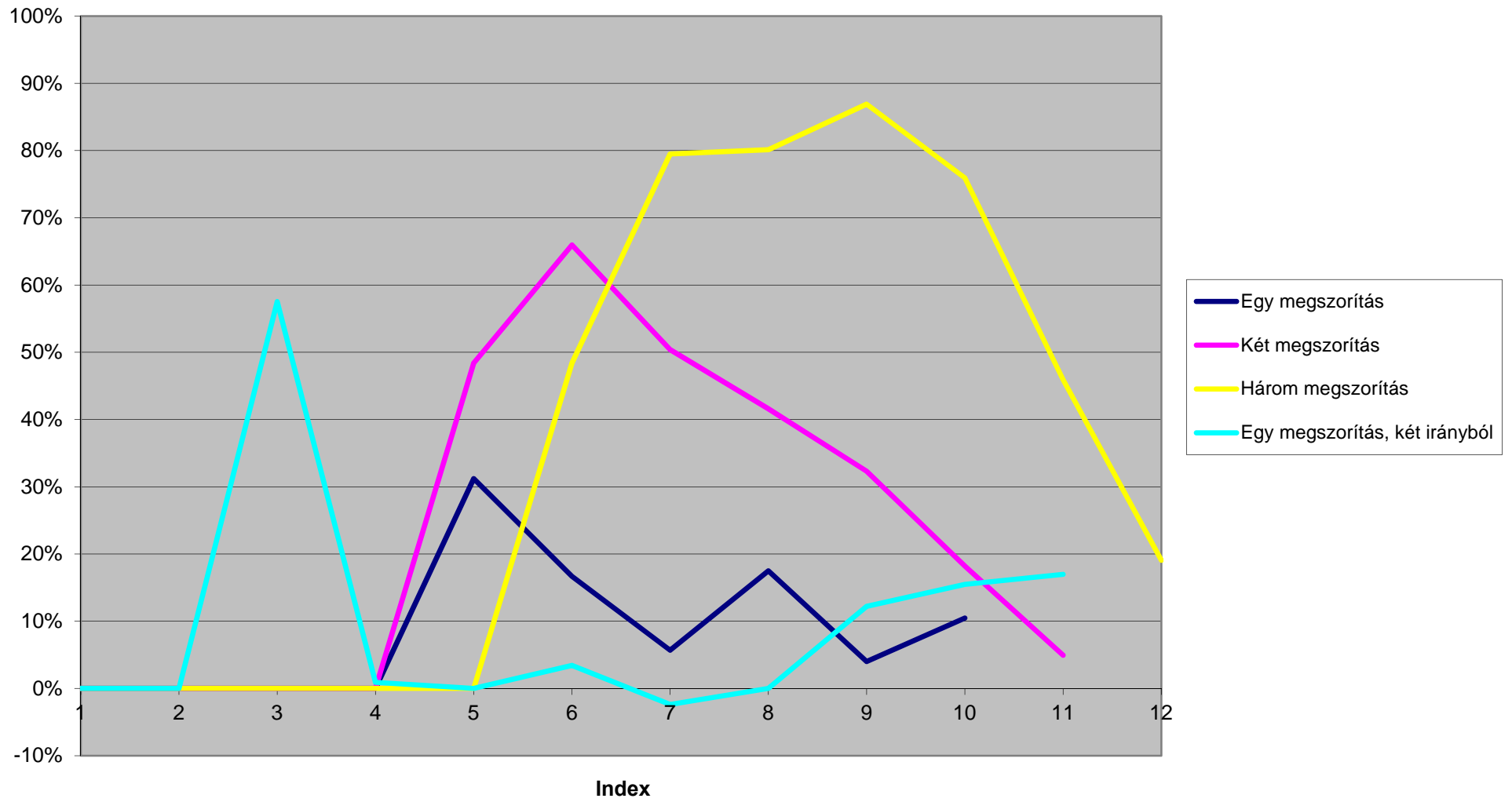
<http://www.ars.usda.gov/Services/docs.htm?docid=20959>

NDB NO	SHRT DESC	WATER	ENERG KCAL	PROTEIN	LIPID TOT	ASH	CARBOHYDRT	FIBER TD	SUGAR TOT
1004	CHEESE,BLUE	42,41	353	21,4	28,74	5,11	2,34	0	0,5
1009	CHEESE,CHEDDAR	36,75	403	24,9	33,14	3,93	1,28	0	0,52
1015	CHEESE,COTTAGE,LOWFAT,2% MILKFAT	80,69	86	11,83	2,45	1,36	3,66	0	3,67
1020	CHEESE,FONTINA	37,92	389	25,6	31,14	3,79	1,55	0	1,55
1025	CHEESE,MONTEREY	41,01	373	24,48	30,28	3,55	0,68	0	0,5
1030	CHEESE,MUENSTER	41,77	368	23,41	30,04	3,66	1,12	0	1,12
1037	CHEESE,RICOTTA,PART SKIM MILK	74,41	138	11,39	7,91	1,15	5,14	0	0,31
1043	CHEESE,PAST PROCESS,PIMENTO	39,08	375	22,13	31,2	5,84	1,73	0,1	0,62
1047	CHEESE FD,PAST PROCESS,SWISS	43,67	323	21,92	24,14	5,77	4,5	0	
1001	BUTTER,WITH SALT	15,87	717	0,85	81,11	2,11	0,06	0	0,06
1002	BUTTER,WHIPPED,WITH SALT	15,87	717	0,85	81,11	2,11	0,06	0	0,06

# Adatforrás, 46 féle összetevő

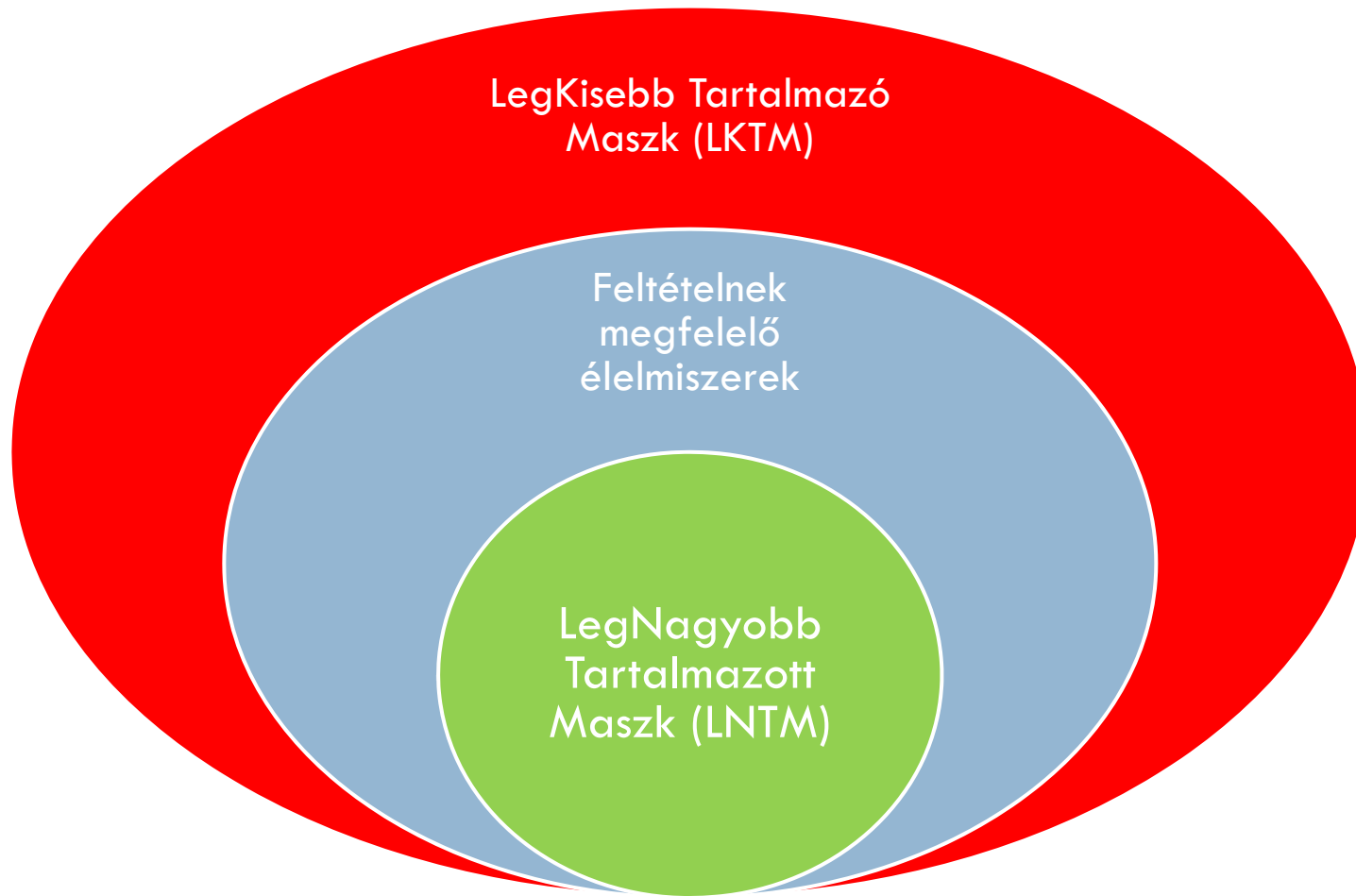
- <http://www.ars.usda.gov/Services/docs.htm?docid=20959>
- WATER ENERG\_KCAL PROTEIN LIPID\_TOT ASH CARBOHYDRT FIBER\_TD  
SUGAR\_TOT CALCIUM IRON MAGNESIUM PHOSPHOROUS POTASSIUM  
SODIUM ZINC COPPER MANGANESE SELENIUM VIT\_C THIAMIN RIBOFLAVIN  
NIACIN PANTO\_ACID VIT\_B6 FOLATE\_TOT FOLIC\_ACID FOOD\_FOLATE  
FOLATE\_DFE CHOLINE\_TOT VIT\_B12 VIT\_A\_IU VIT\_A\_RAE RETINOL  
ALPHA\_CAROT BETA\_CAROT BETA\_CRYPT LYCOPENE LUT\_ZEA VIT\_E  
VIT\_D\_MCG VIT\_D\_IU VIT\_K FA\_SAT FA\_MONO FA\_POLY CHOLESTRL

### Az indexes keresés előnye az egyes méréseknél





# IMEE további alkalmazása



# IMEE további alkalmazása

- Egy szűk halmazt gyorsan lekérdezni (LNTM).
- Ezt a szűk halmazt gyorsan eljuttatni a felhasználóhoz.
- Ha a felhasználó elidőzik a listán, akkor a lista finomítása:
  - ▣ LKTM – LNTM halmaz meghatározása.
  - ▣ Ezen halmazból leválogatni az eredeti feltételeknek megfelelő elemeket.
  - ▣ Leválogatás közben az új elemek folyamatos eljuttatása a felhasználóhoz.

KÖSZÖNÖM A FIGYELMET!

