

# Augmenting search using a semantic visual graph

Edwin de Jonge

Olav ten Bosch

*Statistics Netherlands*



# Contents

- **The use of SVG at Statistics Netherlands**
- **StatLine in a Nutshell**
- **Searching StatLine**
- **Searching using a Semantic Visual Graph**
- **Conclusions**



# Example1: Economic Compass

CBS - Publications - Economic monitor - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

← Vorige → Zoeken Favorieten


Adres <http://www.cbs.nl/en/publications/articles/macro-economics/business-cycle/economic-indicators/compass/compass.htm> Ga naar Koppelingen >>

**Statistics Netherlands**  
Home Site-map Search Nederlands

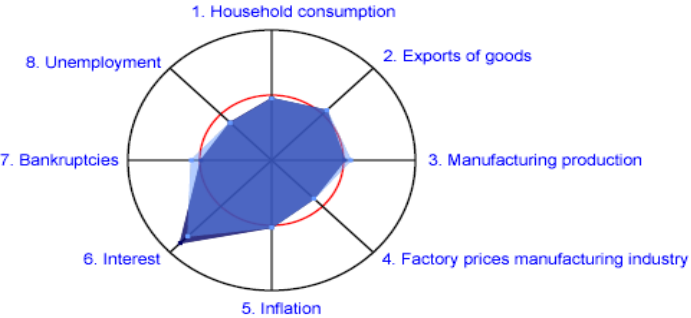
Figures  
Publications  
Service  
Standards  
Organisation

Publications - **Economic monitor**

**Economic compass**

[Explanatory note](#) 

**The larger the blue area, the better the economy is doing.**



1. Household consumption  
2. Exports of goods  
3. Manufacturing production  
4. Factory prices manufacturing industry  
5. Inflation  
6. Interest  
7. Bankruptcies  
8. Unemployment

Picture at the end of July 2005 Picture on 10 August 2005

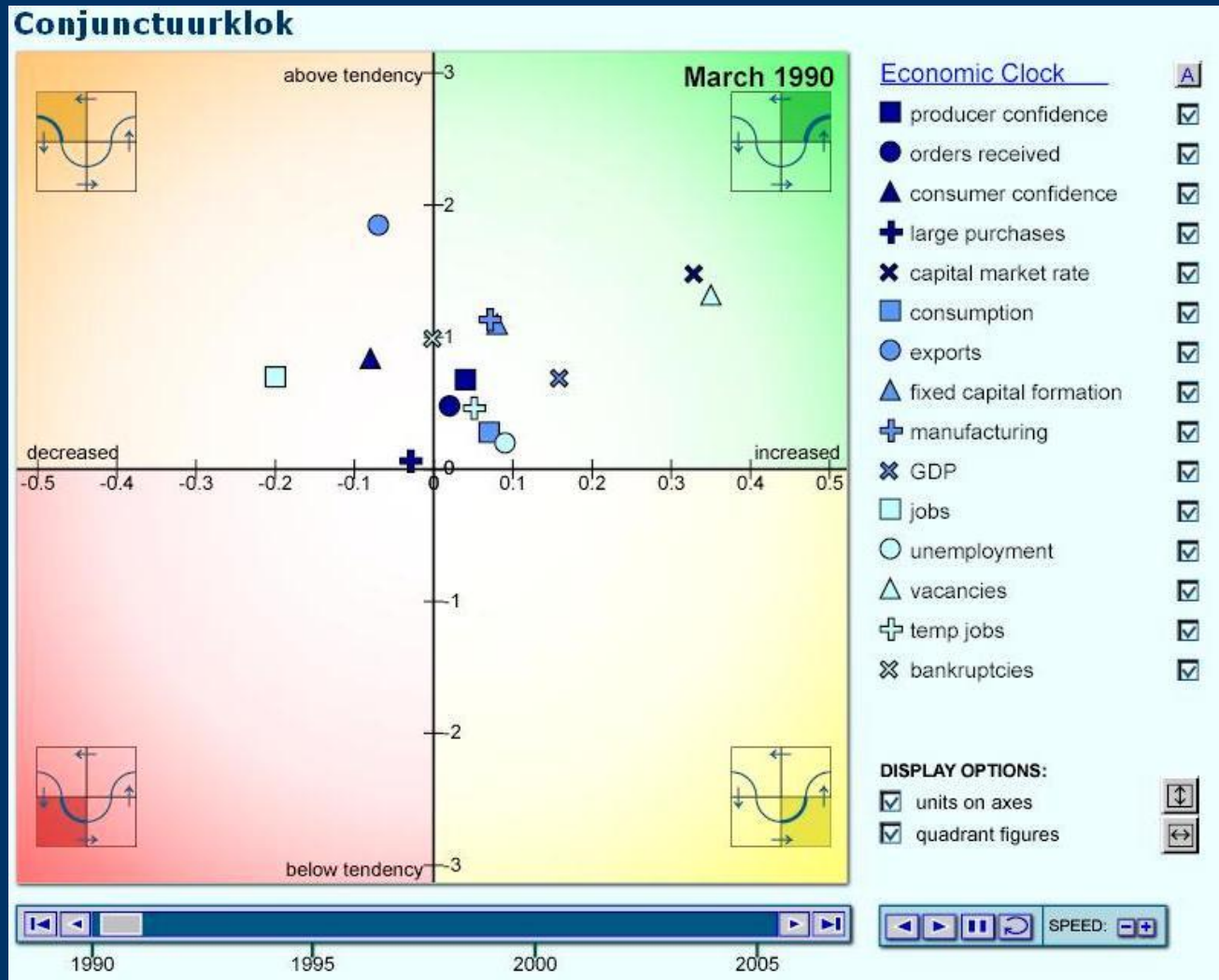
Red circle represents the average of 2003-2004.

© Statistics Netherlands, Voorburg/Heerlen, 2005  
Information may be quoted provided the source is stated accurately and clearly.  
Reproduction for own/internal use is permitted.  
[Privacy statement](#) | [Disclaimer](#)

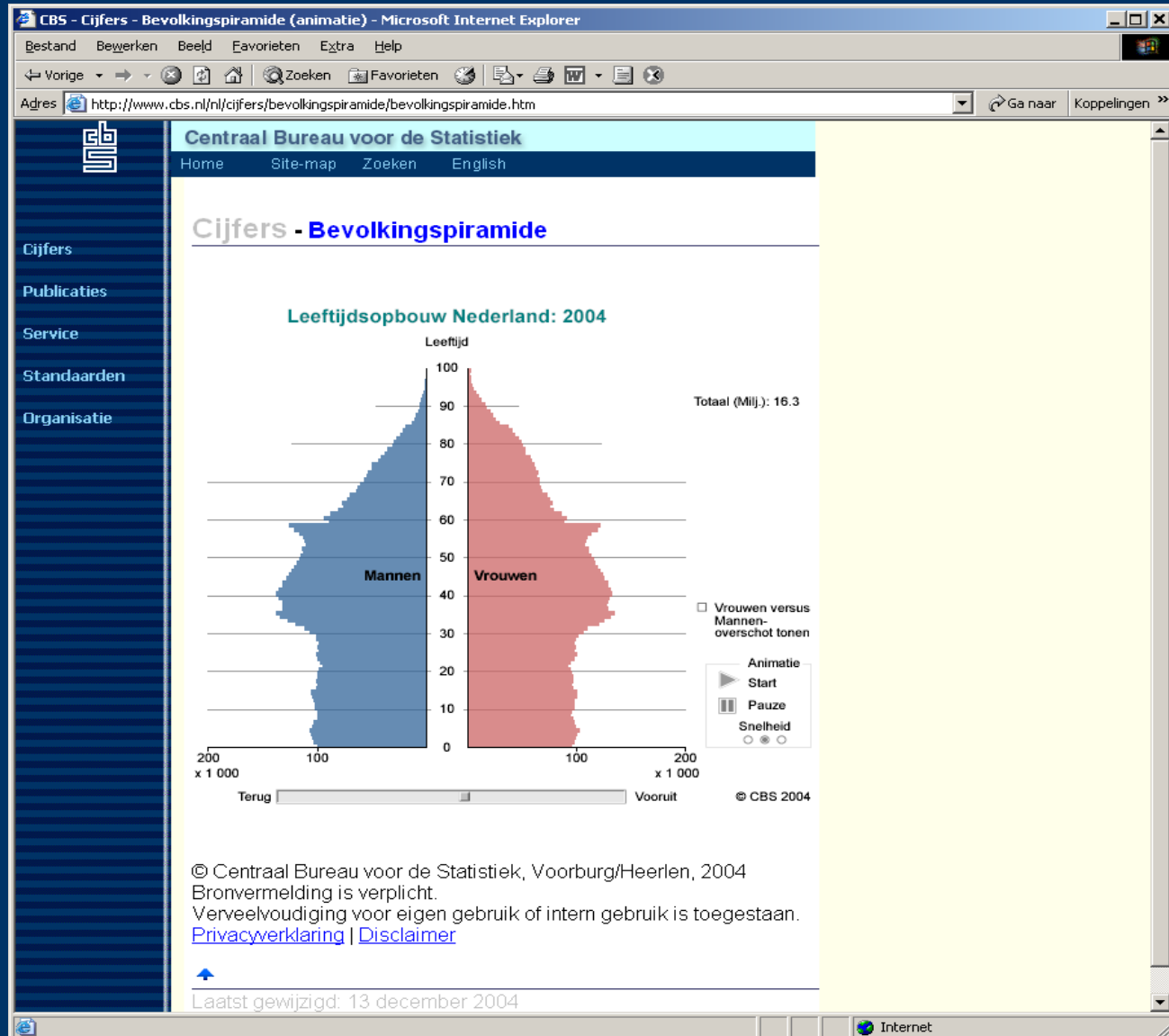
Gereed Internet



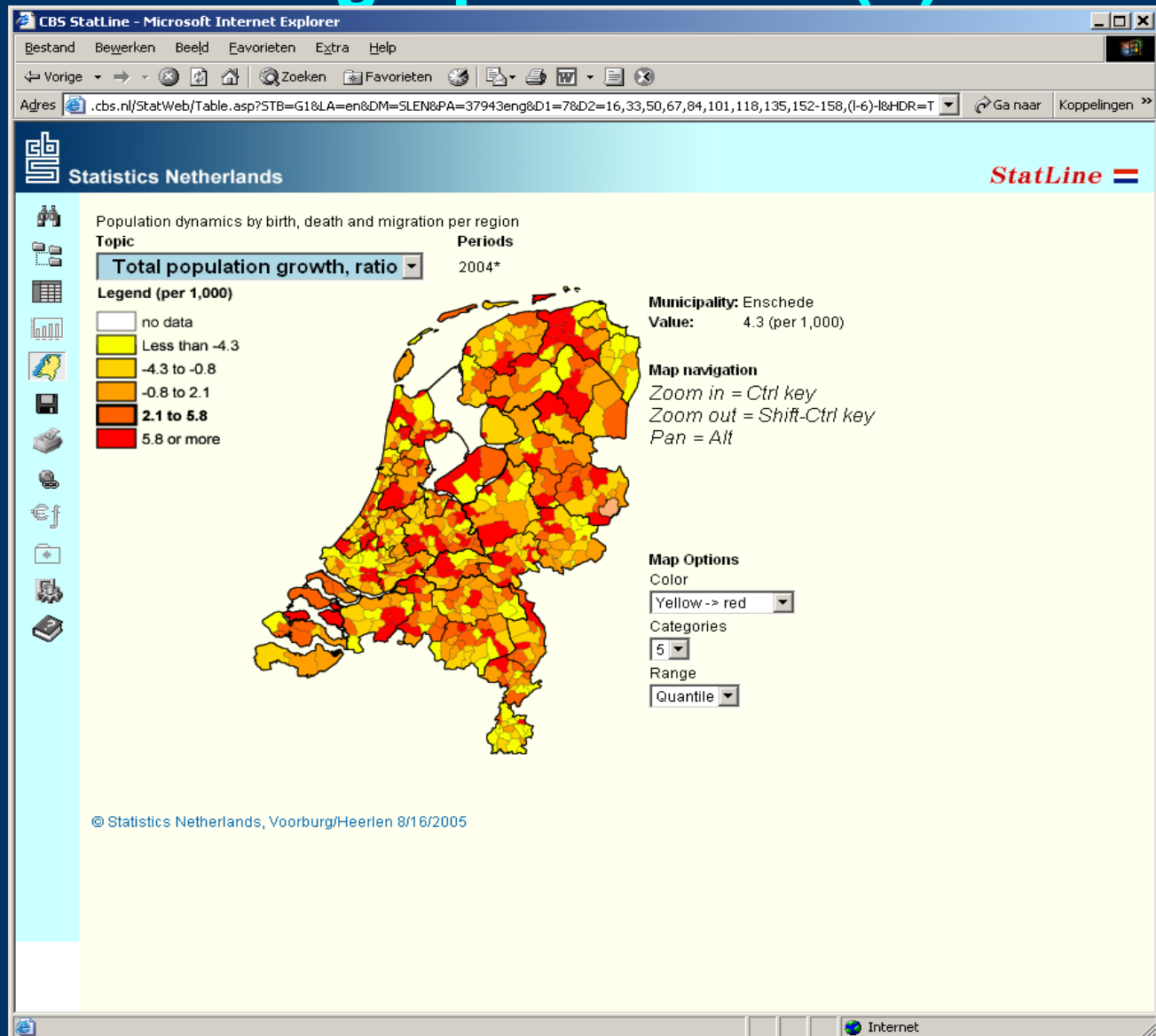
# Example2: Economic “Clock”



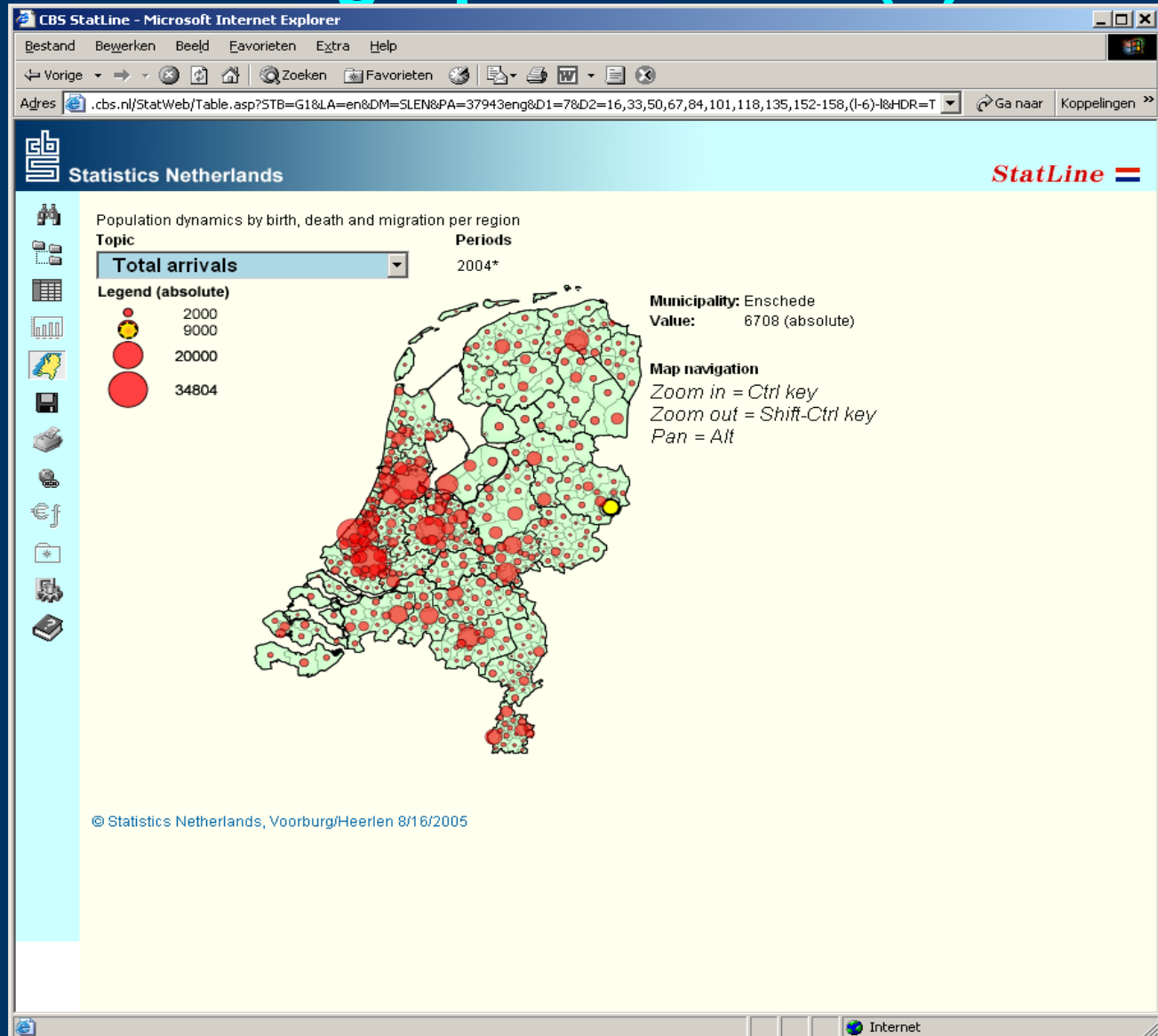
# Example3: Population Pyramid



# Ex.4: Geographical Info (1)



# Ex.5: Geographical Info. (2)



# StatLine in a Nutshell (1)

## - StatLine:

- Statistical database freely available on the internet (<http://statline.cbs.nl>)
- Over 2 billion data cells with 100 Million statistical facts
- Multidimensional cubes with hierarchical dimensions organized by theme
- Search facility
- Tables, charts and maps
- DUAL = refer to any table, map or chart in StatLine within one URL
  - (example: give the 10 most recent statistical facts on subject x)





# StatLine in a Nutshell (2)

The screenshot shows the CBS StatLine website in Microsoft Internet Explorer. The browser's address bar displays the URL: <http://statline.cbs.nl/StatWeb/Start.asp?p=Search/Search&LA=EN&DM=SLEN>. The website header includes the Statistics Netherlands logo and the StatLine logo with the Dutch flag.

### Consumer price index (CPI) all households, 2000=100

CPI: 00000 Total expenditure

Subjects	CPI	Derived CPI
Periods	2000=100	
2000	100.0	100.0
2001	104.2	103.1
2002	107.6	106.4
2003	109.9	108.4
2004	111.2	109.4

© Statistics Netherlands, Voorburg/Heerlen 4/10/2005

### Statline Webselector

**Main groups**

- Macro economy, money and capital markets
  - National accounts
  - Economic data
  - Money and capital markets
  - Prices
    - Consumer prices
      - CPI; all households (2000=100) (selected)
      - CPI; househ. of wage earners

**Selection**

CPI; all households (2000=100) (2590)

Subjects (2) | CPI (259) | Periods (5)

- CPI
  - All categories (259)
  - All groups
    - Divisions (15)
    - Groups (40)
    - Classes (95)

Consumer price indices all households. Index figures for consumer goods. (base year 2000=100). 2000-2004 ; January 2000-March 2005. Changed on 7 April 2005.

Show Data | Help | Refresh

Java Applet Window



# Searching in StatLine (1)

## - Search engine:

- Uses an index on a multidimensional search space of 2 billion cells
- Full text search as well as keyword search
- Preselects data in small sets of tables
- Provides default selections for dimensions without hits
- Uses a weighting scheme for metadata
- Offers web service access




# Searching in StatLine (2)

CBS StatLine - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

Vorige → Zoeken Favorieten Ga naar Koppelingen »

Adres <http://statline.cbs.nl/StatWeb/Start.asp?p=Search/Search&LA=EN&DM=SLEN>

Statistics Netherlands StatLine 

## Search

Search in StatLine for:  period:

Using:

Example of a freetext query: population Rotterdam, period: 2001  
Example of an advanced query: (agriculture or fishery) period:(2000 or 2001)

**7 tables have been created for you:**

1. Found: *inflation*  
**Table created from publication 'Harmonised price index figures' (36 cells)**  
Harmonised price index figures (1996=100); The Netherlands, Eurozone, European Union and year-on-year change, the Netherlands, Eurozone and European Union; 1996-2004; January 1996-July 2005. Changed on 3 August 2005. Frequency: 24 times a year.  
[Alter the selected variables for this table](#)
2. Found: *inflation*  
**Table created from publication 'CPI; all households (2000=100)' (12 cells)**  
Consumer price indices all households, index figures Consumer goods (base year 2000=100), 2000-2004; January 2000-July 2005. Changed on 3 August 2005. Frequency: Monthly  
[Alter the selected variables for this table](#)
3. Found: *inflation*  
**Table created from publication 'CPI; all househ., details (1995=100)' (36 cells)**  
Consumer price indices all households. Index figures Consumer goods. 1996-2002; January 1996-December 2002. Changed on 9 september 2003. Frequency: discontinued on 1 January 2003.

Gereed Internet



# Searching in StatLine (3)

**Full text search engine works reasonably well**

**But:**

**Finding data in a 2 billion search space still difficult**

- **Synonyms/jargon cause less retrieval**
  - **Consumer Price Index?**
  - **Inflation!**
- **Homonyms cause too much retrieval**
  - **“banen” (Dutch)**
    - Jobs, labor statistics
    - Swimming lanes, recreational statistics



# Searching in StatLine (4)

## Fundamental keyword search problem:

- Assumes user knows exact and specific keywords
- Many users use unspecific search words:
  - Economy, Environment
- Due to?:
  - Formulation, user tries to find right question (synonym related)
  - Exploration, user explores interesting direction



# Semantic Visual Graph (1)

## Add semantics to search

- Give semantic feedback to user

## Explicit semantic model:

- Xml topic maps (XTM)
- Supports Scope/Context

## Show semantic neighborhood current search words:

- Show Semantic Visual Graph
- Every topic is a node
- Every association is an edge



# Semantic Visual Graph (2)

The screenshot shows a Microsoft Internet Explorer window titled "WebForm1 - Microsoft Internet Explorer". The address bar contains "http://rdv5t/viewer/webform1.aspx". The search bar at the top contains the text "student" and has buttons for "Zoeken" and "Wis".

The main content area is divided into two sections:

- Search Interface (Zoeken):** A yellow box containing search options. It includes a "Zoek in" field with "StatLine" selected, a "periode:" field, and a "Met behulp van:" dropdown menu set to "Vrij zoeken". Below this, there are example sentences and a "Zoeken" button.
- Search Results:** A list of results. The first result is "1. Gevonden: student" with a table icon and the text "Tabel gemaakt uit publicatie 'Open universiteit; studenten' (204 cellen)". The second result is "2. Gevonden: student" with a table icon and the text "Tabel gemaakt uit publicatie 'WO voltijd; rendement en uitval' (1560 cellen)".

On the right side of the browser window, a semantic graph is displayed. It features a central node labeled "student". Lines connect this central node to several other nodes: "persoon", "cohortjaar", "eerstejaarsstudent", and "vooropleiding".



# Semantic Visual Graph (2)

The screenshot shows a Microsoft Internet Explorer browser window titled "WebForm1 - Microsoft Internet Explorer". The address bar contains "http://rdv5t/viewer/webform1.aspx". The search bar at the top contains the word "persoon" and a "Zoeken" button. Below the search bar is a "Wis" button. The main content area is divided into two parts. On the left is a search results panel titled "Zoeken" with a search icon. It contains a search form with "persoon" entered in the "Zoek in StatLine" field. Below the form are options for "Met behulp van" (set to "Vrij zoeken") and a "Zoeken" button. There are also example sentences and a list of search results. The first result is "1. Gevonden: *persoon*" followed by a link to a table titled "Tabel gemaakt uit publicatie 'Gebruik medische voorzieningen' (3516 cellen)". The second result is "2. Gevonden: *persoon*" followed by a link to "Tabel gemaakt uit publicatie 'Regionale inkomensverdeling'". On the right side of the browser window is a semantic visual graph. The central node is "persoon". Lines radiate from this central node to various related terms: "nationaliteit", "weduwe persoc", "student", "vrouw", "allochtoon persoon", "inkomen", "autochtoon p", "gescheiden p", "huishouden", "woonplaats", "verhuisd persoon", "man", "leeftijd", "kerkelijke gezindte", "docent", "geslacht", "burgelijke staat", "ongetrouwd persoc", "hoofdbewor", and "ethniciteit".





# Semantic Visual Graph (2)

The screenshot shows a Microsoft Internet Explorer window titled "WebForm1 - Microsoft Internet Explorer". The address bar contains "http://rdv5t/viewer/webform1.aspx". The search bar at the top has "inkomen" entered and "Zoeken" and "Wis" buttons. Below the search bar is a "Zoeken" section with a search form. The search form has "inkomen" in the "StatLine" field and "periode:" followed by an empty field. The "Met behulp van:" dropdown is set to "Vrij zoeken". Below the search form, there are two example sentences: "Voorbeeld van vrij zoeken: verhuizingen buitenland, periode: 2001" and "Voorbeeld van geavanceerd zoeken: (vakens of mest) periode: (2000 of 2001)". Below this, it says "Er zijn 100 tabellen voor u samengesteld:". The first result is "1. Gevonden: *inkomen*" followed by a link to "Tabel gemaakt uit publicatie 'Regionale inkomensverdeling 1999' (116 cellen)". The description for this table is: "Besteedbaar inkomen en inkomensverdelingen van personen en huishoudens Per gemeente (1-1-2000), COROP-gebied, provincie, landsdeel, stads- gewesten en grootstedelijke agglomeraties, 1999 Gewijzigd op: 2 juli 2004 Verschijningsfrequentie: eenmalig". Below the description is a link "Kies zelf de variabelen voor de te bouwen tabel". The second result is "2. Gevonden: *inkomen*" followed by a link to "Tabel gemaakt uit publicatie 'Regionale inkomensverdeling 2000' (116 cellen)". To the right of the search results is a semantic graph with nodes "persoon", "particuliere huishouden", "inkomen", and "bevolking". Lines connect "persoon" to "particuliere huishouden", "particuliere huishouden" to "inkomen", and "inkomen" to "bevolking".

inkomen Zoeken Wis

**Zoeken**

Zoek in StatLine  periode:

naar:

Met behulp van:  Zoeken

Van:

Voorbeeld van vrij zoeken: verhuizingen buitenland, periode: 2001  
Voorbeeld van geavanceerd zoeken: (vakens of mest) periode: (2000 of 2001)

**Er zijn 100 tabellen voor u samengesteld:**

1. Gevonden: *inkomen*

**Tabel gemaakt uit publicatie 'Regionale inkomensverdeling 1999'** (116 cellen)

Besteedbaar inkomen en inkomensverdelingen van personen en huishoudens Per gemeente (1-1-2000), COROP-gebied, provincie, landsdeel, stads- gewesten en grootstedelijke agglomeraties, 1999 Gewijzigd op: 2 juli 2004 Verschijningsfrequentie: eenmalig

**Kies zelf de variabelen voor de te bouwen tabel**

2. Gevonden: *inkomen*

**Tabel gemaakt uit publicatie 'Regionale inkomensverdeling 2000'** (116 cellen)

persoon

particuliere huishouden

inkomen

bevolking



# Semantic Visual Graph (2)

The screenshot shows a Microsoft Internet Explorer window titled "WebForm1 - Microsoft Internet Explorer". The address bar contains "http://rdv5t/viewer/webform1.aspx". The main content area is divided into two parts:

**Search Interface (Left):**

- A search box containing the text "bevolking".
- Buttons for "Zoeken" and "Wis".
- A section titled "Zoeken" with a search icon.
- Fields for "Zoek in StatLine" (containing "bevolking") and "periode:".
- A dropdown menu for "Met behulp van" set to "Vrij zoeken" and a "Zoeken" button.
- Example text: "Voorbeeld van vrij zoeken: verhuizingen buitenland, periode: 2001" and "Voorbeeld van geavanceerd zoeken: (varkens of mest) periode: (2000 of 2001)".
- Text: "Er zijn 100 tabellen voor u samengesteld:".
- Search results:
  1. Gevonden: *bevolking*
    - Icon of a table.
    - Link: **Tabel gemaakt uit publicatie 'Werkloosheid en beroepsbevolking'** (84 cellen)
    - Text: "Werkzame en werkloze beroepsbevolking, niet beroepsbevolking, werkloosheidspercentage naar seizoenscorrectie, geslacht en leeftijd. 3-maandsperiode dec. 99/feb. 00 - apr. 05/jun. 05, jaartot. 2000-2004. Gewijzigd op 21 juli 2005. Verschijningsfrequentie: per maand."
    - Icon of a table.
    - Text: "Kies zelf de variabelen voor de te bouwen tabel"
  2. Gevonden: *bevolking*
    - Icon of a table.
    - Link: **Tabel gemaakt uit publicatie 'Historie bevolking'** (4000 cellen)

**Semantic Visual Graph (Right):**

- A central node labeled "bevolking".
- Lines radiating from "bevolking" to other nodes: "bevolkingsdichtheid", "vergrijzing", "beroepsbevolking", "inkomen", and "inwoner".

The browser's taskbar at the bottom shows the Windows logo, a search icon, and a status bar with "Onbekende zone (Gemengd)".



# Semantic Visual Graph

## Technically:

- Prototype in cooperation with Delft University
- Tools:
- ASP.NET (IIS)
- C#
- Javascript
- StatLine Webservice
- XTM .Net engine



# SVG features (to be) used

## -textPath:

- naming associations

## - Animation:

- animating from central topic to another

## - Javascript interaction:

- update of graph with javascript callback functions

## - Ajax

- for immediate feedback in visual graph when typing in search box



# Conclusions

**Augmenting search with Semantic Visual Graph based on SVG looks promising**

**But...:**

- Search improvement must be quantified
- Usability tests have to be performed
- Describing Semantics of a statistical domain in a Semantic model should be explored further

