Flash Fill and Equation Discovery

Craps Jeroen en De Groote Tom

27 oktober 2014

Overzicht

- Ons doel
- Example-based learning
- Spreadsheet tabel transformaties
- Equation discovery
- Conclusie

Overzicht

- **Ons doel**
- Example-based learning
- Spreadsheet tabel transformaties
- Equation discovery
- Conclusie

Ons doel

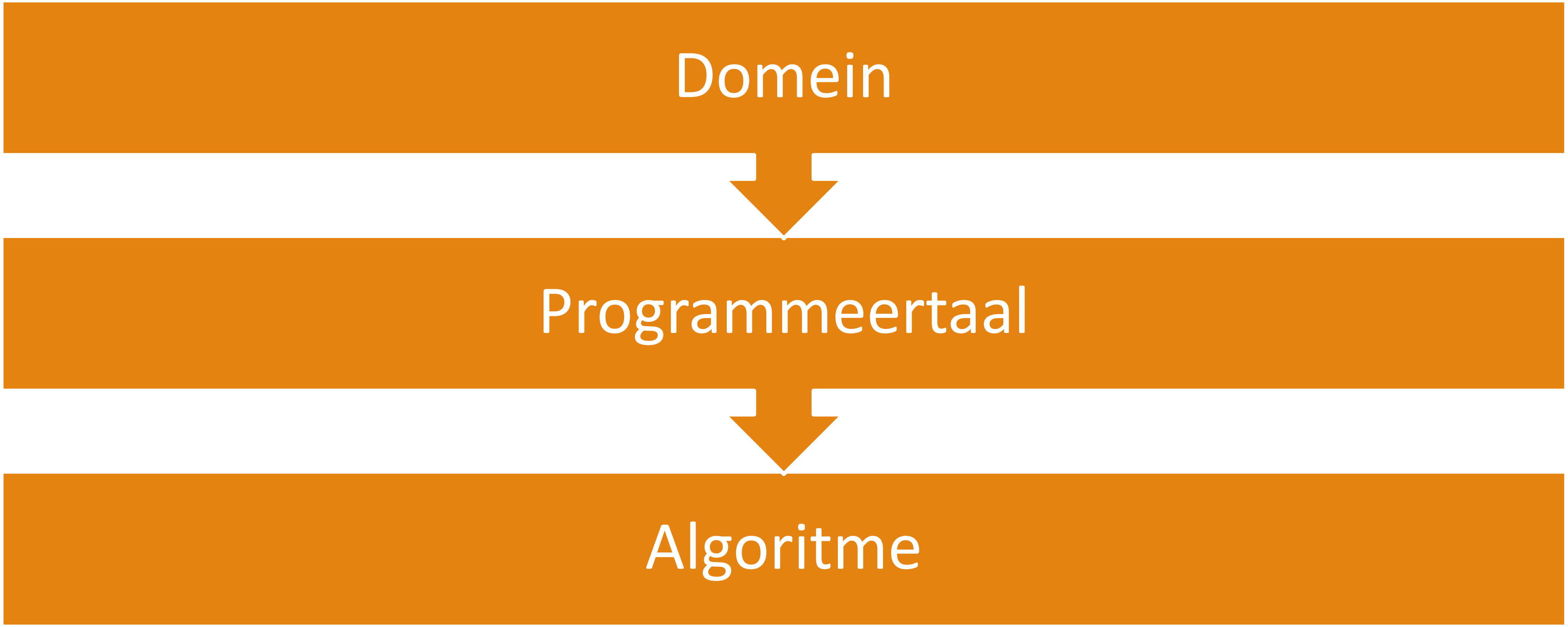
- Flash fill voor getallen
  - Equation discovery voor nodig
- Voorbeeld

X1	X2	X3	Y
1	2	2	9
4	3	3	
1	8	2	
8	9	2	
2	2	2	

→

X1	X2	X3	Y
1	2	2	9
4	3	3	343
1	8	2	81
8	9	2	289
2	2	2	16

End-user programming



Domein

Programmeertaal

Algoritme

Overzicht

- Ons doel
- **Example-based learning** [Gulwani, CACM 14]
- Spreadsheet tabel transformaties
- Equation discovery
- Conlusie

Vergelijking: soorten problemen

PROCEDUREEL

- Beslissingsmethode
- Memoriseren en toepassen

CONCEPTUEEL

- Geen beslissingsmethode
- Creatief denken
- Bewijs en Constructie problemen

Probleem generatie

- Genereren adhv voorbeeld problemen
- Voordelen
  - Voorkomen spieken
  - Copyright problemen vermijden

Oplossing generatie

- Automatisch generen oplossing
- Voordelen
  - Voorbeeld oplossing
  - Oplossing op basis van deeloplossing
  - Hints

Feedback generatie

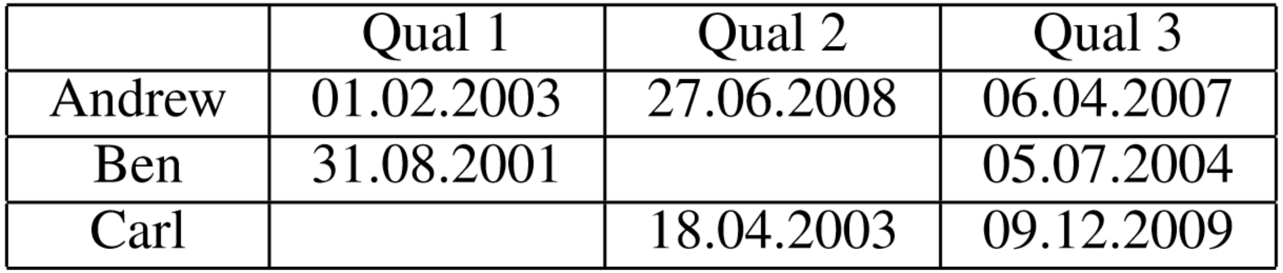
- Controleren correctheid oplossing
  - Waarom incorrect
  - Waar en hoe fout verbeteren (hint)
- Iedereen gelijke beoordeling
- Handmatig analyseren fouten tijdrovend

Overzicht

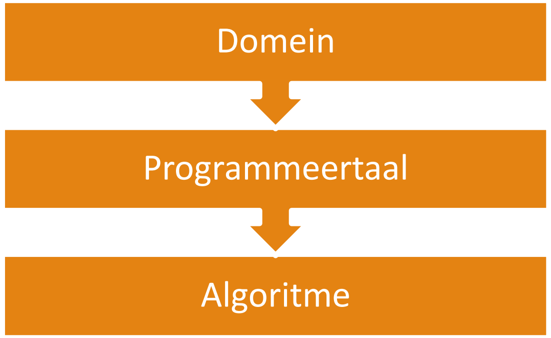
- Ons doel
- Example-based learning
- **Spreadsheet tabel transformaties** [Harris en Gulwani, CACM 11]
- Equation discovery
- Conclusie

Concept oplossing generatie

Voorbeeld input tabel:



	Qual 1	Qual 2	Qual 3
Andrew	01.02.2003	27.06.2008	06.04.2007
Ben	31.08.2001		05.07.2004
Carl		18.04.2003	09.12.2009

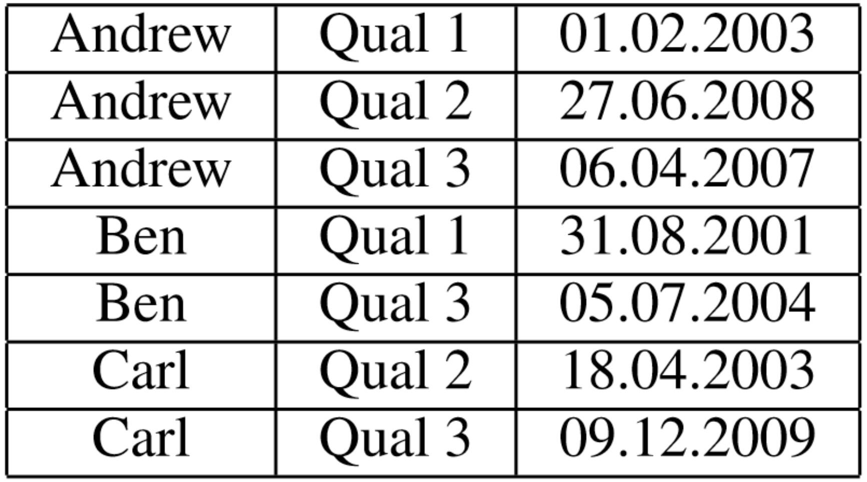


Domein

Programmeertaal

Algoritme

Voorbeeld output tabel:



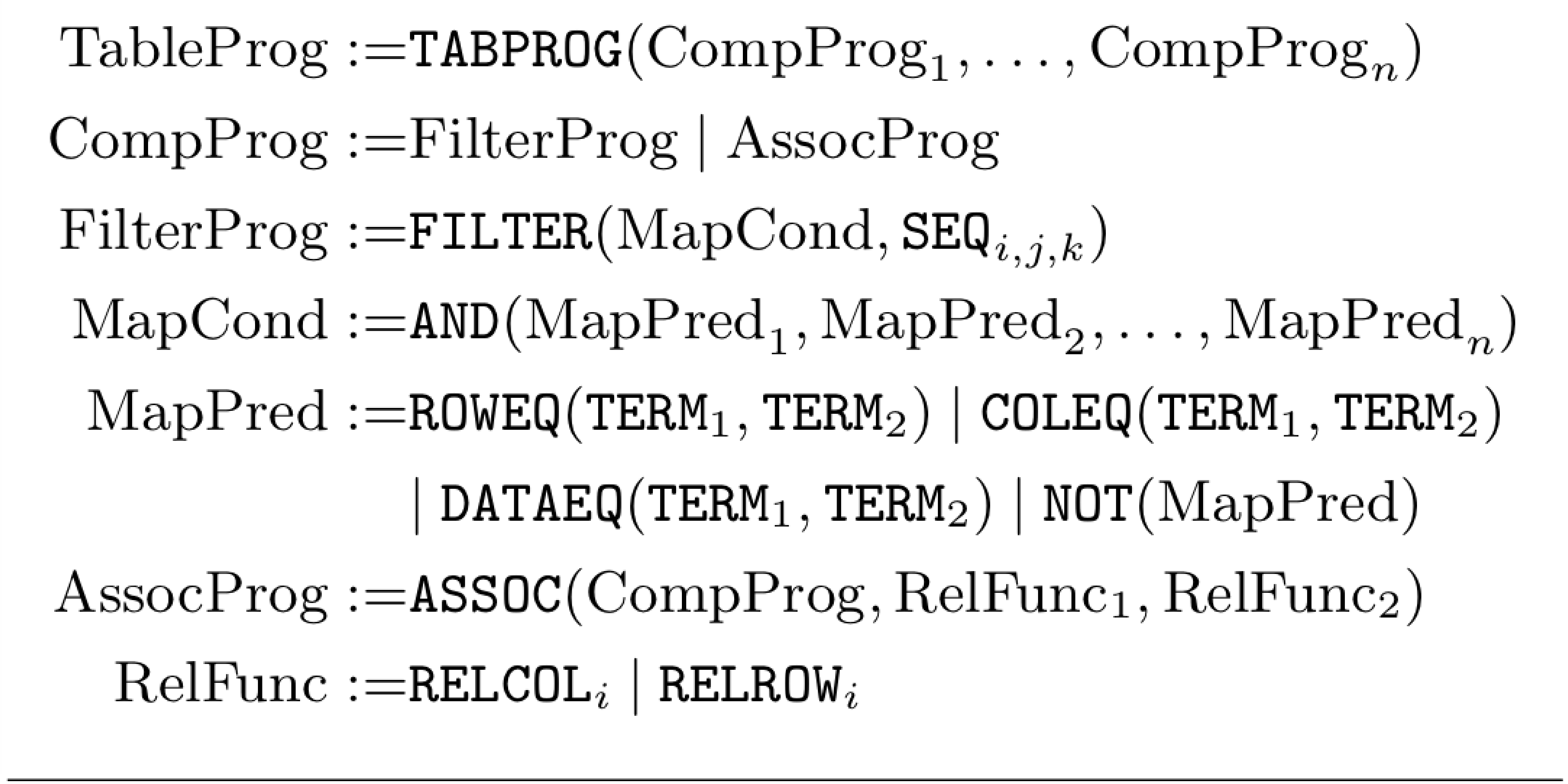
Andrew	Qual 1	01.02.2003
Andrew	Qual 2	27.06.2008
Andrew	Qual 3	06.04.2007
Ben	Qual 1	31.08.2001
Ben	Qual 3	05.07.2004
Carl	Qual 2	18.04.2003
Carl	Qual 3	09.12.2009

[Harris en Gulwani, CACM 11]

Tabel transformaties

1. Programmeertaal: TableProg

Grammatica

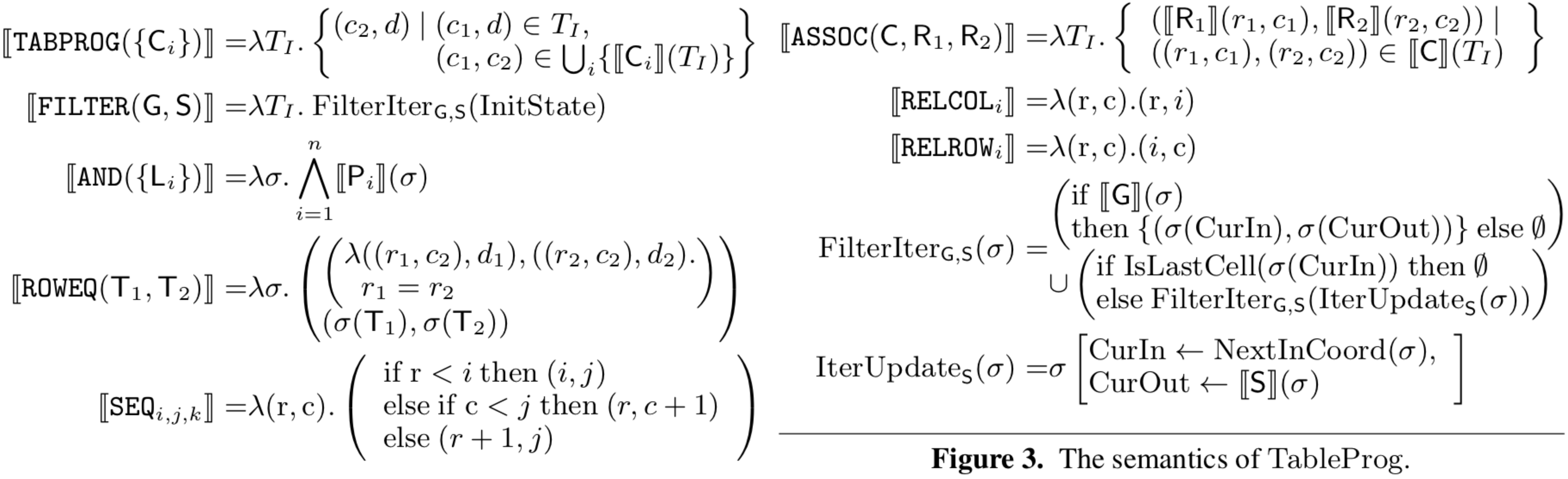


$$
\begin{aligned}
\text{TableProg} :=& \texttt{TABPROG}(\text{CompProg}_1, \ldots, \text{CompProg}_n)\\
\text{CompProg} :=& \text{FilterProg} \mid \text{AssocProg}\\
\text{FilterProg} :=& \texttt{FILTER}(\text{MapCond}, \texttt{SEQ}_{i,j,k})\\
\text{MapCond} :=& \texttt{AND}(\text{MapPred}_1, \text{MapPred}_2, \ldots, \text{MapPred}_n)\\
\text{MapPred} :=& \texttt{ROWEQ}(\texttt{TERM}_1, \texttt{TERM}_2) \mid \texttt{COLEQ}(\texttt{TERM}_1, \texttt{TERM}_2)\\
& \mid \texttt{DATAEQ}(\texttt{TERM}_1, \texttt{TERM}_2) \mid \texttt{NOT}(\text{MapPred})\\
\text{AssocProg} :=& \texttt{ASSOC}(\text{CompProg}, \text{RelFunc}_1, \text{RelFunc}_2)\\
\text{RelFunc} :=& \texttt{RELCOL}_i \mid \texttt{RELROW}_i
\end{aligned}
$$

Figure 2. The syntax of TableProg.

[Harris en Gulwani, CACM 11]

Domein kennis



$$[\![\texttt{TABPROG}(\{\mathsf{C}_i\})]\!] = \lambda T_I.\left\{\begin{array}{r}(c_2,d)\mid(c_1,d)\in T_I,\\ (c_1,c_2)\in\bigcup_i\{[\![\mathsf{C}_i]\!](T_I)\}\end{array}\right\}$$

$$[\![\texttt{FILTER}(\mathsf{G},\mathsf{S})]\!] = \lambda T_I.\ \mathrm{FilterIter}_{\mathsf{G},\mathsf{S}}(\mathrm{InitState})$$

$$[\![\texttt{AND}(\{\mathsf{L}_i\})]\!] = \lambda\sigma.\ \bigwedge_{i=1}^{n}[\![\mathsf{P}_i]\!](\sigma)$$

$$[\![\texttt{ROWEQ}(\mathsf{T}_1,\mathsf{T}_2)]\!] = \lambda\sigma.\left(\begin{array}{l}\left(\begin{array}{l}\lambda((r_1,c_2),d_1),((r_2,c_2),d_2).\\ \quad r_1=r_2\end{array}\right)\\ (\sigma(\mathsf{T}_1),\sigma(\mathsf{T}_2))\end{array}\right)$$

$$[\![\texttt{SEQ}_{i,j,k}]\!] = \lambda(\mathrm{r},\mathrm{c}).\left(\begin{array}{l}\text{if } \mathrm{r}<i \text{ then } (i,j)\\ \text{else if } \mathrm{c}<j\text{ then }(r,c+1)\\ \text{else }(r+1,j)\end{array}\right)$$

$$[\![\texttt{ASSOC}(\mathsf{C},\mathsf{R}_1,\mathsf{R}_2)]\!] = \lambda T_I.\left\{\begin{array}{l}([\![\mathsf{R}_1]\!](r_1,c_1),[\![\mathsf{R}_2]\!](r_2,c_2))\mid\\ ((r_1,c_1),(r_2,c_2))\in[\![\mathsf{C}]\!](T_I)\end{array}\right\}$$

$$[\![\texttt{RELCOL}_i]\!] = \lambda(\mathrm{r},\mathrm{c}).(\mathrm{r},i)$$

$$[\![\texttt{RELROW}_i]\!] = \lambda(\mathrm{r},\mathrm{c}).(i,\mathrm{c})$$

$$\mathrm{FilterIter}_{\mathsf{G},\mathsf{S}}(\sigma) = \begin{array}{l}\left(\begin{array}{l}\text{if }[\![\mathsf{G}]\!](\sigma)\\ \text{then }\{(\sigma(\mathrm{CurIn}),\sigma(\mathrm{CurOut}))\}\text{ else }\emptyset\end{array}\right)\\ \cup\left(\begin{array}{l}\text{if IsLastCell}(\sigma(\mathrm{CurIn}))\text{ then }\emptyset\\ \text{else FilterIter}_{\mathsf{G},\mathsf{S}}(\mathrm{IterUpdate}_{\mathsf{S}}(\sigma))\end{array}\right)\end{array}$$

$$\mathrm{IterUpdate}_{\mathsf{S}}(\sigma) = \sigma\left[\begin{array}{l}\mathrm{CurIn}\leftarrow\mathrm{NextInCoord}(\sigma),\\ \mathrm{CurOut}\leftarrow[\![\mathsf{S}]\!](\sigma)\end{array}\right]$$

Figure 3. The semantics of TableProg.

[Harris en Gulwani, CACM 11]

Tabel transformaties

1. TableProg
2. Algoritme: ProgFromEx
3. Experimenten

Overzicht

- Ons doel
- Example-based learning
- Spreadsheet tabel transformaties
- **Equation discovery** [Todorovski, LNAI 07]
- Conclusie

Definitie

*“**Equation discovery** is a machine learning task that deals with the problem of learning **quantitative laws and models**, expressed in the form of equations, in collections of measured **numeric data**.”*
– Encyclopedia of machine learning

Kennis integreren

- 3 manieren
  - Expliciete integratie
  - Impliciete integratie
  - Inductive bias
- Inductive language bias

Context vrije grammatica (1)

*“A **context free grammar** consists of a finite set of variables , each of them representing a subclass of subexpressions or phrases in the language represented by the grammar.”*

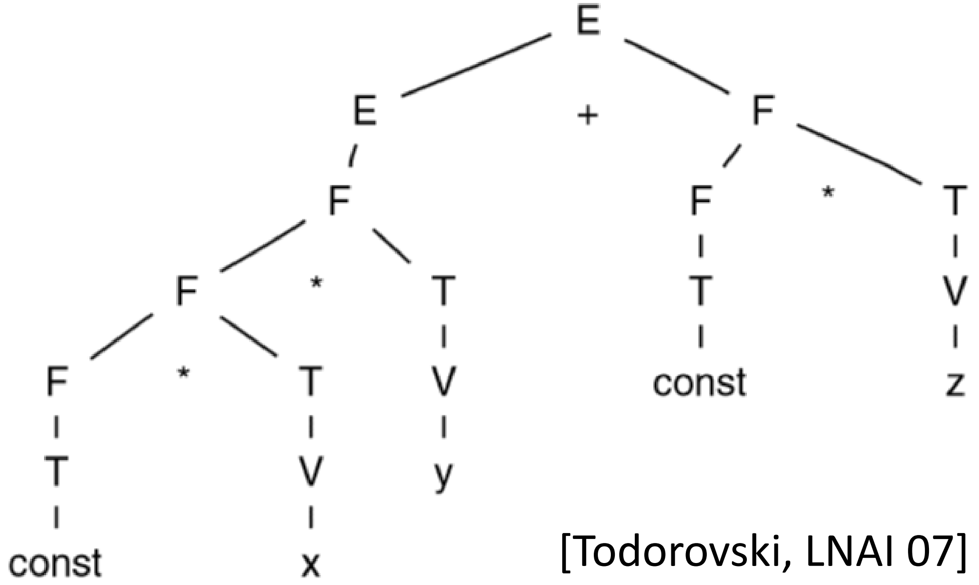
- Todorovski & Džeroski

Context vrije grammatica (2)

- Voorbeeld context vrije grammatica

$$\begin{aligned} E &\rightarrow E + F \mid E - F \\ F &\rightarrow F * T \mid F/T \\ T &\rightarrow (E) \mid V \mid const \end{aligned}$$

- Parse tree
  - Controleren expressie tot taal behoort
  - Expressies generen die tot taal behoren



[Todorovski, LNAI 07]

Overzicht

- Ons doel
- Example-based learning
- Spreadsheet tabel transformaties
- Equation discovery
- **Conclusie**

Conclusie

- Ons doel: Flash Fill voor getallen
  - Domein kennis: equation discovery
  - Grammatica
  - Genereren van verschillende opties

Papers

- **Spreadsheet Table Transformations from Examples**, CACM, W.R. Harris, S. Gulwani
- **Example-Based Learning in Computer-Aided STEM Education**, CACM 2014, Sumit Gulwani
- **Integrating domain knowledge in equation discovery**, LNCS 4660, Springer Berlin Heidelberg, 2007, Todorovski, Ljupčo, and Sašo Džeroski
- **Equation Discovery**, Enclyclopedia of Machine Learning

Vragen?