

# FlashQuation: Vergelijkingen vinden aan de hand van voorbeelden

Jeroen Craps & Tom De Groote

## Introductie

### Probleemgebied:

De slechte gebruikerservaring bij het zoeken naar een passende vergelijking.

### Onderzoeksvraag:

Is het mogelijk om in beperkte tijd een vergelijking te vinden in de door gebruikers gedefinieerde set van getallen?

### Hypothese:

Het is mogelijk om een vergelijking te vinden in beperkte tijd over een gegeven set getallen.

## Methoden

### STAPPEN:

1. Eenmalig de bewerkingsboom bereken aan de hand van CFG.
2. Verwijderen redundante knopen.  
→ Optimale bewerkingsboom.
3. Voorbeeld wordt gegeven.
4. Efficiënt zoeken van vergelijkingen die voldoen aan het gegeven voorbeeld.

### Hypothese 1:

Er is een significant verschil tussen het aantal knooppunten van een optimale boom en van de volledige boom.

### Hypothese 2:

Er kunnen meer unieke oplossingen gevonden worden in beperkte tijd met behulp van de optimale boom.

Voor de zoekruimte te verbreden worden constanten toegevoegd aan de CFG.

### Hypothese 3:

Door middel van deze constanten neemt de algemene oplossingsgraad toe.

Om de oplossingsgraad te verhogen worden geen overbodige vergelijkingen uitgerekend. (vb.  $T_0 * T_1 / T_1 \sim T_0$ )

### Hypothese 4:

Het efficiënt evalueren verhoogt de oplossingsgraad ten opzichte van de brute-force manier.

## Contextvrije grammatica (CFG)

### Productieregels van de vorm:

$E \rightarrow T$  (R1)

$E \rightarrow E O T$  (R2)

$T \rightarrow 1..9 | K1 | K2 | K3$

met E een niet-terminaal symbool, T een terminaal symbool en O een operatie.

### Voorbeeld:

E

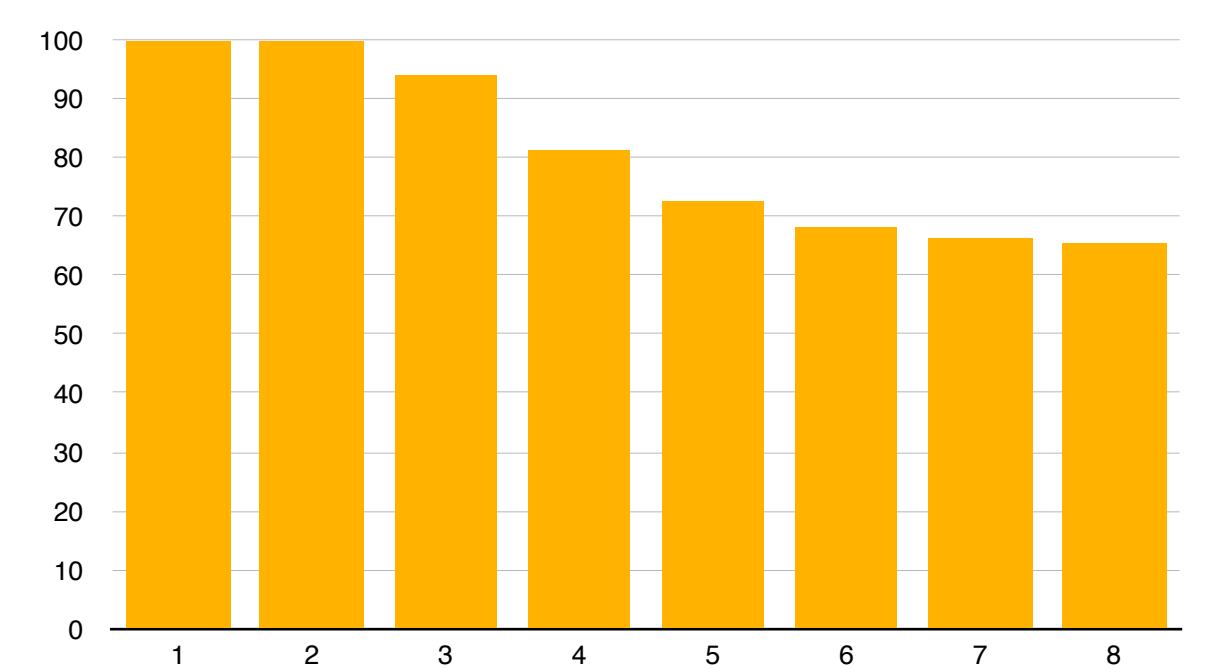
→ E O T

→ T O T

T wordt bij het evalueren ingevuld door gegeven waarde of gewichten.

## Resultaten

### Pruned vs Non-pruned:



Procentuele weergave van het aantal knooppunten van de optimale boom in vergelijking met de volledige boom.

## Conclusie

Met behulp van contextvrije grammatica kan er in sommige gevallen een vergelijking gevonden in een volledig willekeurige set van getallen. Indien de vergelijking niet volledig willekeurig is, dan neemt de kans op het vinden van een vergelijking enorm toe.

### Toekomstig onderzoek:

- Wat is de optimale afweging tussen oplossingsgraad en snelheid?
- Is er een heuristiek die de complexiteit significant verbeterd?
- Etc...

## Referenties:

- Spreadsheet Table Transformations from Examples, CACM, W.R. Harris, S. Gulwani
- Example-Based Learning in Computer-Aided STEM Education, CACM 2014, Sumit Gulwani
- Equation Discovery, Encyclopedia of Machine Learning



Begeleiders: Prof. L. De Raedt  
Postdoc. A. Kimmig

Contact: jeroen.craps@student.kuleuven.be  
tom.degroote@student.kuleuven.be

KU LEUVEN