

# Generation of Hypergraphs from the N-Best Parsing of 2D-Probabilistic Context-Free Grammars for Mathematical Expression Recognition

E. Noya, J. A. Sánchez, J.M. Benedí

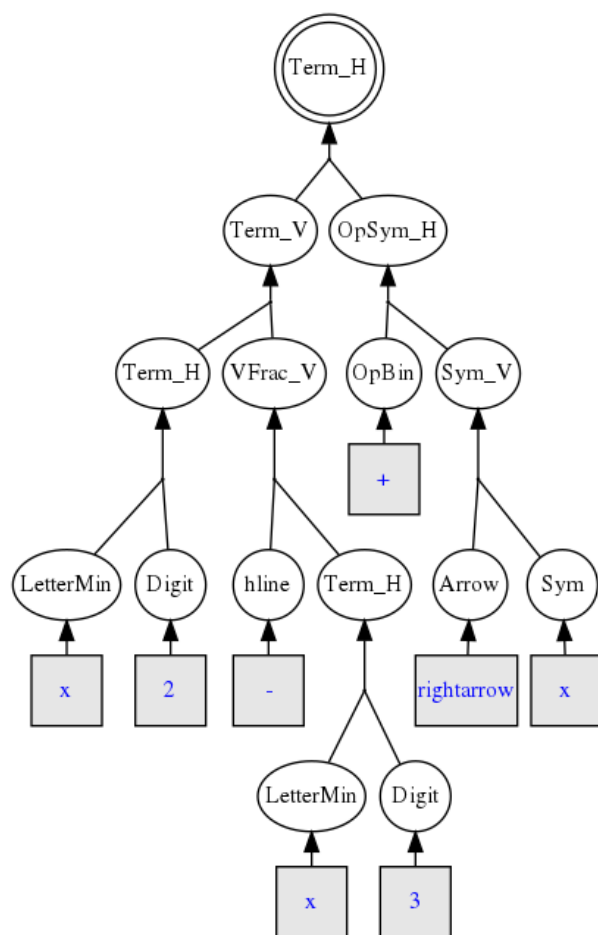
*PRHLT, Universitat Politècnica de València, Spain*

`{noya, jandreu, jmbenedi}@prhlt.upv.es`

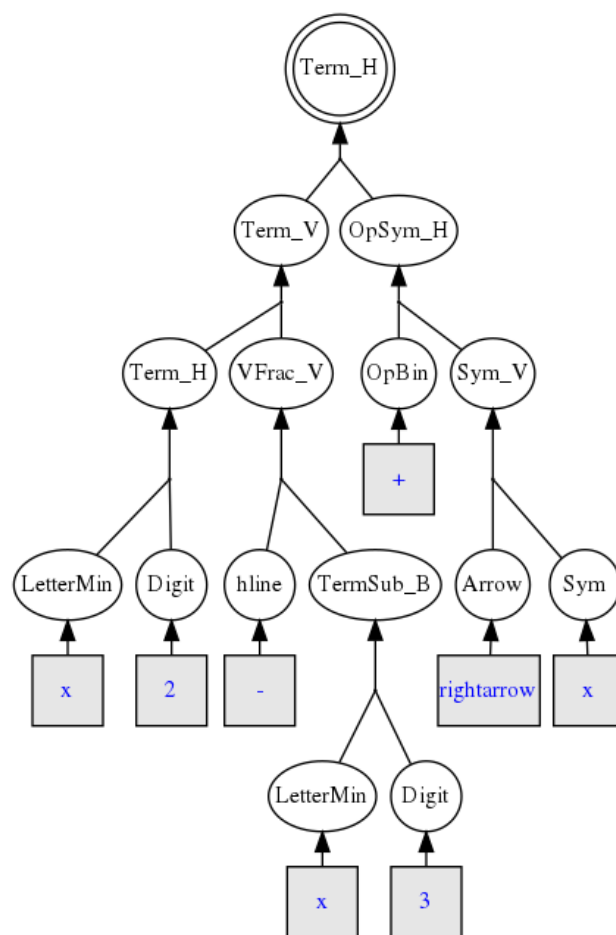
# Introduction

- A new approach for searching math expressions (ME) in large collections of printed document images has been introduced.
- This approach does not require any kind of segmentation, and
- It does not need to have a complete and error-free transcription of the images.
- To reduce the search time, a two-phase solution is proposed.
  - a) **Off-line phase**, the posterior probabilities of MEs are calculated from hypergraphs derived from the Mathematical Expression Recognition process.
  - b) **On-line phase**, these posteriors are used for indexing and searching for MEs in the collection.
- In this paper, we focus on the first off-line phase, and the main contributions include:
  1. the computation of the n-best parse trees from a 2D-PCFGs,
  2. the generation of hypergraphs from the n-best parse trees.

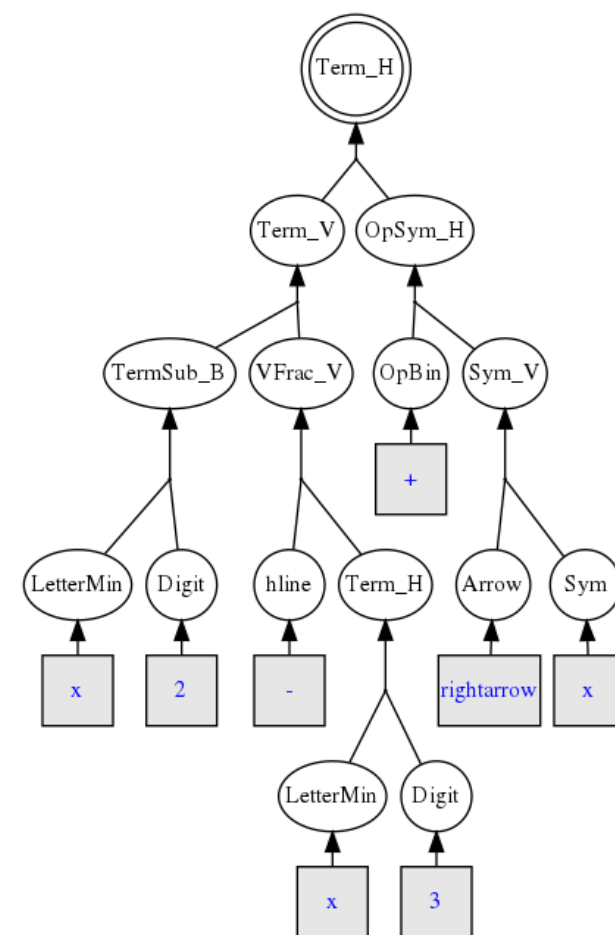
# N-best parse trees from a 2D-PCFGs



$$\frac{x2}{x3} + \vec{x}$$

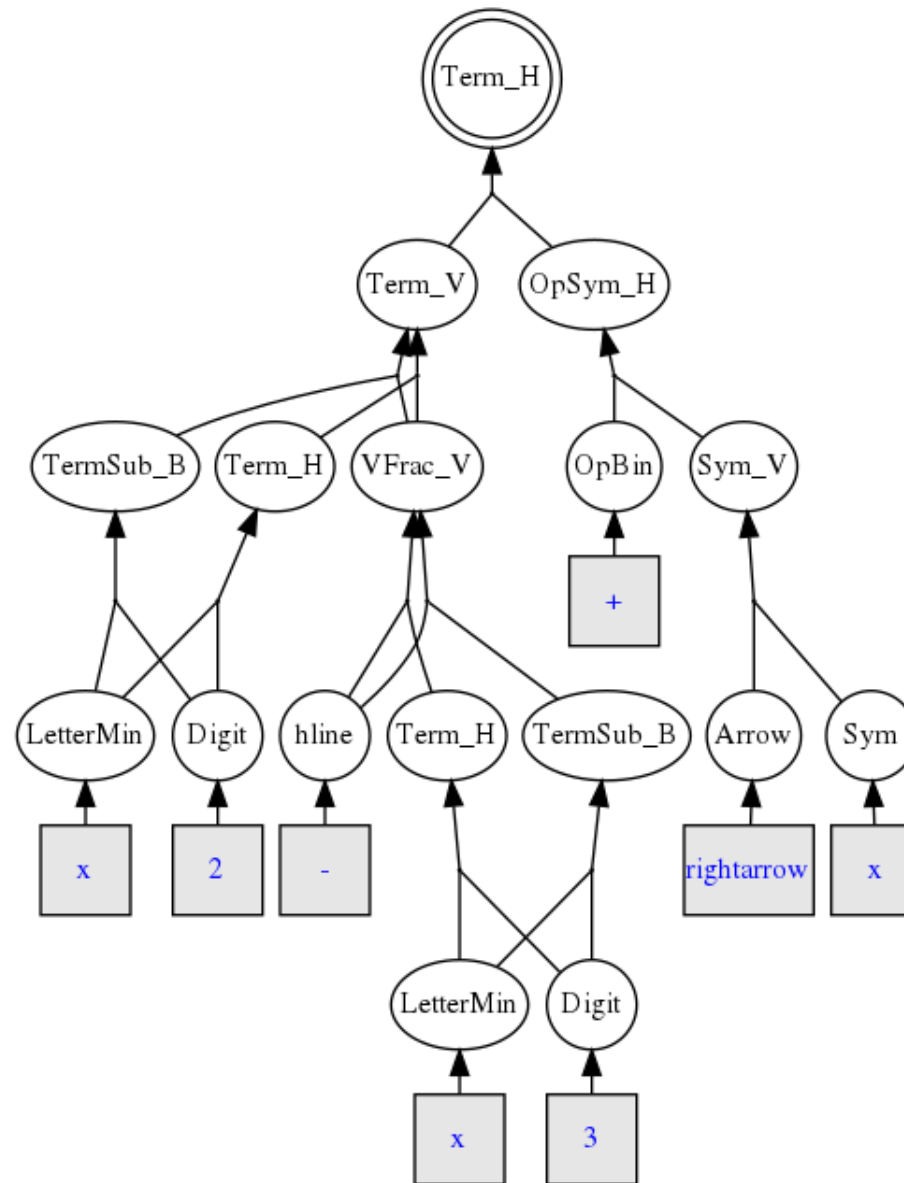


$$\frac{x2}{x3} + \vec{x}$$



$$\frac{x2}{x3} + \vec{x}$$

# Hyperforest from N-best parse trees



## Conclusions

- A proposal for generating hypergraphs from the N-Best parsing of 2D-PCFGs for ME recognition has been presented.
- A formal framework for the development of inference algorithms (*in-side* and *outside*) and normalization strategies of hypergraphs has been also presented.