The Structure of Mathematical Expressions

An arXiv Case Study

Deyan Ginev and Bruce R. Miller

National Institute of Standards and Techonology

March 21, 2012

Contents

Chapter 1

Introduction

In this study, we survey the notational diversity of present-day mathematical expressions, in order to uncover their linguistic phenomena. A practical motivation for this study is to provide a foundation for determining the boundary between syntactic and semantic phenomena in said expressions, from the perspective of language modeling. The ultimate goal of this project is to construct a grammar of mathematical expressions, which captures all relevant syntactic properties established in this study, and allows for the semantic analysis necessary to model and observe the semantic relationships.

1.1 Motivation

We want to enable machine-reading of formulas, in order to provide a variety of user-assistance services, such as semantic search, text-to-speech synthesis, semantic interactions (definition lookup), as well as computer algebra support ("evaluate subexpressions on demand") and ultimately computer verification ("does that proof step really hold?").[1] EdN:1

1.2 Related Resources

Notation census, beginnings of study are in Deyan's thesis, Naproche and FMathL have examples, but no real systematic study.[2] EdN:2

1.3 Experimental Setup

The primary corpus on which we base this investigation is the Cornel pre-print archive "arXiv"[3], consisting of over 700,000 articles in 37 scientific subfields. EdN:3

[1] EdNote: expand
[2] EdNote: expand
[3] EdNote: cite here

Train1	Differential Geometry http://arxmliv.kwarc.info/files/9609/dg-ga.9609012
Train2	Quantum Physics http://arxmliv.kwarc.info/files/0910/0910.5733/
Train3	High Energy Physics - Theory http://arxmliv.kwarc.info/files/9407/hep-th.9407125/
Train4	Commutative Algebra http://arxmliv.kwarc.info/files/0809/0809.4873/
Train5	Statistics Theory http://arxmliv.kwarc.info/files/0905/0905.1486/
Train6	General Relativity and Quantum Cosmology http://arxmliv.kwarc.info/files/0807/0807.2507/
Train7	Cosmology and Extragalactic Astrophysics http://arxmliv.kwarc.info/files/0908/0908.2548
Train8	Exactly Solvable and Integrable Systems http://arxmliv.kwarc.info/files/0905/0905.2033
Train9	Geometric Topology http://arxmliv.kwarc.info/files/0809/0809.4477
Train10	Algebraic Geometry http://arxmliv.kwarc.info/files/0704/0704.0537

Table 1.1: Sandbox of Ten Random arXiv Papers from Diverse Scientific Subfields

arXiv Sandbox

EdN:4 [4]

As a secondary resource, we we will also consult entry-level literature on highschool mathematics, in order to exhibit basic phenomena, as well as to demonstrate phenomena
EdN:5 apriori known to the authors.[5]

[4] EdNote: Say that, on the arXiv front, we first start with the train sandbox from Deyan's thesis
[5] EdNote: Wikipedia? PEMDAS?

Chapter 2

A Study of Mathematical Syntax

2.1 Basics

Foundations

6 7 8

EdN:6
EdN:7
EdN:8

High School

9 10

EdN:9
EdN:10

2.2 Discrete math

Set Theoretic Notations

11 12

EdN:11
EdN:12

Logical Operators

13

EdN:13

[6] EdNote: arithmetic, grouping fences and equality
[7] EdNote: basic relations and orderings
[8] EdNote: arithmetic and algebraic sequences?
[9] EdNote: geometry here, otherwise a separate geometry subsection
[10] EdNote: trigonometry, complex and rational numbers
[11] EdNote: elementhood, inclusions, set constructors, overloaded arith ops
[12] EdNote: also maps : domains -¿ codomains, xRy notations
[13] EdNote: classic logic, HOL, type theories

Combinatorics

EdN:14 [14] [15]
EdN:15

Number Theory

EdN:16 [16] [17] [18] [19]
EdN:17

Graph Theory

EdN:18
EdN:19 [20] [21] [22]
EdN:20
EdN:21

Algebra

EdN:22 [23] [24] [25] [26]
EdN:23
EdN:24

Functions Theory

EdN:25 [27]
EdN:26
EdN:27

2.3 Continuous math

Calculus

EdN:28 [28]

Probability

EdN:29 [29] [30]
EdN:30

[14] EdNote: Infinite sums
[15] EdNote: binomials, combinations, permutations,
[16] EdNote: modulo modifiers
[17] EdNote: tuples
[18] EdNote: divisibility notations $a \mid b$ and b/a
[19] EdNote: DLMF sneaky notations
[20] EdNote: edge and vertex notations
[21] EdNote: incidence and adjacency notations
[22] EdNote: Wiki is very nice: http://en.wikipedia.org/wiki/Glossary_of_graph_theory
[23] EdNote: vectors
[24] EdNote: maps and complements
[25] EdNote: groups
[26] EdNote: lattices
[27] EdNote: talk about associativity of application and composition, ";" and "$\circ$" as notation variants, discuss complex examples
[28] EdNote: differentials, integrals, limits, remember brownian motion integral notations!
[29] EdNote: Bayes formula with multiple denotations of P
[30] EdNote: Various conditional and joint probability notations

Interval Notation and Arithmetic

[31] EdN:31

Topology

[32] EdN:32

2.4 Other fields

Quantum Physics

[33] [34] ⋮ EdN:33
EdN:34

[31] EdNote: introduce interval notations, then move to interval arithmetic
[32] EdNote: manifold constructors and notations
[33] EdNote: Bra-ket notation
[34] EdNote: computer science, biology, chemistry...

Chapter 3

Discussion

Chapter 4

Conclusion