The Structure of Mathematical Expressions

An arXiv *Case Study*

Deyan Ginev and Bruce R. Miller

National Institute of Standards and Techonology

March 16, 2012

Contents

Chapter 1

Introduction

In this study, we survey the notational diversity of present-day mathematical expressions, in order to uncover its linguistic phenomena.

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

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 apriori known to the authors

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

Chapter 2

A Study of Mathematical Syntax

2.1 Basics

Foundations

[4] [5] [6]

EdN:4
EdN:5
EdN:6

High School

[7] [8]

EdN:7
EdN:8

2.2 Discrete math

Set Theoretic Notations

[9] [10]

EdN:9
EdN:10

Logical Operators

[11]

EdN:11

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

Combinatorics

EdN:12 [12] [13]
EdN:13

Number Theory

EdN:14 [14] [15] [16] [17]
EdN:15
EdN:16
Graph Theory

EdN:17 [18] [19] [20]
EdN:18
EdN:19
EdN:20
Algebra

EdN:21 [21] [22] [23] [24]
EdN:22
EdN:23
Functions Theory

EdN:24 [25]
EdN:25

2.3 Continuous math

Calculus

EdN:26 [26]

Probability

EdN:27 [27] [28]
EdN:28

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

Interval Notation and Arithmetic

[29] EdN:29

Topology

[30] EdN:30

2.4 Other fields

Quantum Physics

[31] [32] ⋮ EdN:31
EdN:32

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

Chapter 3

Discussion

Chapter 4

Conclusion