Erschienen in: BIRNDL 2019 - Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries : Proceedings of the 4th Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2019) co-located with the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019), Paris, France, July 25, 2019 / Chandrasekaran, Muthu Kumar; Mayr, Philipp (Hrsg.). - Aachen : RWTH Aachen, 2019. - (CEUR workshop proceedings ; 2414). - S. 108-115

Towards Formula Concept Discovery and Recognition

Philipp Scharpf[1], Moritz Schubotz[2], Howard S. Cohl[3], and Bela Gipp[2]

[1] Department of Computer and Information Science
University of Konstanz, Germany
first.last@uni-konstanz.de
[2] Department of Information Technology
University of Wuppertal, Germany
last@uni-wuppertal.de
[3] National Institute of Standards and Technology, United States
first.last@nist.gov

Abstract. Citation-based Information Retrieval (IR) methods for scientific documents have proven to be effective in academic disciplines that use many references. In science, technology, engineering, and mathematics (STEM), researchers cite less often but employ mathematical concepts to refer to prior knowledge (Moed et al.). Our long-term goal is to generalize citation-based IR-methods and apply the generalized method to both classical references and mathematical concepts. In this paper, we suggest how mathematical formulae could be cited and define a Formula Concept Retrieval challenge with two subtasks: Formula Concept Discovery (FCD) and Formula Concept Recognition (FCR). While the former aims at the definition and exploration of a Formula Concept that names bundled equivalent representations of a formula, the latter is designed to match a given formula to a prior assigned concept ID. Moreover, we present first Machine Learning based approaches to tackle the FCD and FCR tasks, which we apply to a standardized test-collection (NTCIR arXiv dataset). Our FCD approach yields a recall of 68% for retrieving equivalent representations of frequent formulae, and 72% for extracting the formula name from the surrounding text. FCD and FCR will enable citing formulae within mathematical documents and facilitate semantic search as well as similarity computations for plagiarism detection or document recommender systems.

Keywords: Natural Language Processing · Mathematical Language Processing · Mathematical Information Retrieval · Feature Analysis · Machine Learning

1 Introduction

Documents from Science, Technology, Engineering, and Mathematics (STEM) often contain a significant amount of mathematical formulae. Since they are vital to understanding the content of these documents, semantic search engines or recommender systems need to process and analyze them alongside the text. In information science and technology, the semantics of natural language is typically grasped via conceptualization [25]. In the case of mathematical language, we argue for the introduction of a definition for a mathematical *Formula Concept* as a collection of equivalent formulae with different representations (see [15] for a discussion of the definition difficulties). Once defined, the technical implementation of a Formula Concept can be *Formula Concept Discovery (FCD) and Formula Concept Recognition (FCR)*. The first term (FCD) refers to the exploration of formula concepts by examining a multitude of formula examples from various sources and occurrences. Figure 1 illustrates how the same equation, in this case, the Klein-Gordon equation from Quantum Physics, can be represented in different formats that seem very diverse at first glance but actually represent the same mathematical concept. We will present first implementations of FCD and FCR in the following.

$\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \nabla^2\psi + \left(\frac{m_0 c}{\hbar}\right)^2 \psi = 0$	$u_{tt} + Au + f(u) = 0$
$\partial_{ct}^2 h_n(z,t) - \partial_z^2 h_n(z,t) + \nu_n^2 h_n(z,t) = 0$	$\nabla^a\nabla_a\psi = \mu^2\psi$
$\frac{\hbar^2}{c^2}\frac{\partial^2\psi}{\partial t^2} - \frac{\hbar^2\partial^2\psi}{\partial x^2} = -2i\hbar\frac{\partial\psi}{\partial\tau}$	$-\hbar^2\frac{\partial^2\psi}{\partial t^2} + c^2\hbar^2\nabla^2\psi = m_0^2c^4\psi$
$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} - \frac{2\alpha+a}{c^2}\frac{\partial\phi}{\partial t} - \frac{\alpha^2+a\alpha}{c^2}\phi=0$	$u_{tt} - \Delta u + m^2u + G'(u) = 0$
$\left(\eta^{\mu\nu}\frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x^\nu} - \left(\frac{mc}{\hbar}\right)^2\right)\phi = 0$	$u_{tt} - \Delta u + mu + P'(u) = 0$
$\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \sum_{i=1}^{p} \frac{\partial}{\partial x^i}\frac{\partial}{\partial x^i} - \left(\frac{mc}{\hbar}\right)^2\right)\phi = 0$	$(m > 0, P(u) \geqslant 0)$

Fig. 1: Various representations of the Klein-Gordon equation extracted from physics papers [2], [22], [7], [21], [6], [12], [11], [4], [20].

2 Related Work

Mathematical Information Retrieval (MathIR) addresses the information need in STEM fields by retrieving, processing and analyzing mathematical formulae. Up until now, various formula search engines have been developed, and translations between different markups (LaTeX, Presentation, and Content MathML) and standards elaborated [5]. Since Wikipedia is only semi-structured, Wikidata[4] was launched to provide direct access to specific interlingual facts (RDF[5] triples) and retrieve information systematically. Wikidata is a free and open semantic knowledge-base that can be read and edited by humans and machines [23]. Wikidata stores items with statements and their references. In the case of mathematical knowledge, this includes formulae, e.g., pressure (Q39552) with a defining formula property (P2534) $p = \frac{F}{S}$. To scalably seed information into Wikidata, a Primary Sources tool[6] was introduced, allowing active users to quickly browse through new claims and their references to approve or reject them. The arXiv.org e-Print server [10] makes available free preprints for a large collection of publications from Physics, Mathematics, Computer Science, Economics and more. Many authors provide their LaTeX source code. Both Wikipedia and arXiv articles were extracted as part of the NTCIR MathIR Task [1]. In 2017, the Special Interest group for Math Linguistics (SIGMathLing)[7] was initiated as a forum and resource cooperative for the linguistics of mathematical/technical documents. For Mathematical Language Processing (MLP), the formula parts (operators, identifiers, numbers) have to be annotated using the Mathematical Markup Language (MathML). There are several tools available, most prominently the *LaTeXML* converter[8]. Furthermore, the occurring symbols (variables, constants) need to be disambiguated, i.e., their meaning inferred from the context and semantically annotated. There have been attempts to automatically retrieve the semantics of identifiers from the surrounding text [18]. While Wikipedia articles more commonly contain variable definitions in the text, in general, many paper articles often omit them. This leaves the task of manual annotation inevitable for building machine-interpretable datasets. The NIST Digital Repository of Mathematical Formulae (DRMF) [3] and NIST Digital Library of Mathematical Functions (DLMF) [9] are two examples of maintained high-quality semantic datasets. At this moment, Wikidata contains approximately 3600 items with a "defining formula" property. Moreover, the benchmark MathMLben [17] was created to evaluate tools for mathematical format conversion (from LaTeX to MathML to Computer Algebra Systems), containing approximately 300 formulae from Wikipedia, the arXiV and the DLMF, which were augmented by Wikidata macros [16].

---
[4] `http://www.wikidata.org`
[5] `https://www.w3.org/RDF`
[6] `https://www.wikidata.org/wiki/Wikidata:Primary_sources_tool`
[7] `https://sigmathling.kwarc.info`
[8] `https://dlmf.nist.gov/LaTeXML`

3 Formula Concept Retrieval Challenge

We define as the goal to be eventually able to map all of the various representations of a formula to a unique and open concept ID, e.g., linking all occurrences of the Klein-Gordon equation shown in Figure 1 to the Wikidata item Q868967 [9].

We define two subtasks of the *Formula Concept Retrieval challenge*:

- Formula Concept Discovery (FCD) as a method to find common equivalent representations and a name candidate for a given formula, and
- Formula Concept Recognition (FCR) as the approach to recognize formulae in documents as being instances of prior defined formula concept.

4 Our Approach

In the following, we present our first efforts to implement and evaluate a Formula Concept Discovery (FCD). We approach FCD by retrieving equivalent formulations with different representations (see Figure 2) as well as name candidates from the surrounding text. The initial step is to identify formula candidates which occur most often within a given dataset, assuming that they are potential seeds of popular formula concepts. We first tried formula clustering but discovered that it was not a suitable method for FCD since the number of clusters is a priori unclear and the tested algorithms were not able to group equivalent formulae. Subsequently, we decided to start with a ranking of formula duplicates (with the same LaTeX string), which yielded reasonable results. We employed the NTCIR arXiv dataset [1] which is comprised of 104062 document sections containing over 60 million formulae. We confined our computations to the subject class of astrophysics (680 astro-ph documents), employing a domain expert to semantically evaluate the results. From the duplicate ranking, we selected a formula length range between 10 and 30 characters and restricted our selection to duplicates occurring in at least two different documents. This yielded 3495 formulae. We then manually selected all equations, and discarded all stubs without a right-hand-side, as well as simple variable dependence definitions, such as $x = x(t)$ and $x = y$ or $x = \text{const}$. For the first 50 samples from the duplicate ranking, we retrieved the operators and identifiers from the provided MathML `<mo>` and `<mi>` tags, as well as the surrounding text (words within a window of ± 500 characters around the formula). We encoded both tags using the *TfidfVectorizer* from the Python package *Scikit-learn* [13] and *Doc2Vec* model [8] from the Python package *Gensim* [14]. We then compared the performance of a k-nearest neighbor classifier (Scikit-learn) on the four resulting vector encodings (*math2vec [24]* and *math tf-idf* for the formulae, *semantics2vec* and *semantics tf-idf* for the surrounding text) to retrieve equivalent representations.

[9] https://www.wikidata.org/wiki/Q868967

5 Our Results

Table 1 shows the results of our approach for discovering Formula Concepts. We rank the fetched formulae by the number of duplicates d and also list the number of documents $\hat{d}$ they appear in. The main investigation was to compare the performance of four different encodings in terms of the retrieved number of equivalent representations using the kNN recommendation algorithm provided by Scikit-learn. Calculating the overall success distribution, we discovered that the *math2vec* (e_m) encoding clearly outperforms the others by yielding 71% of the retrieved instances, followed by *semantics tf-idf* ($\hat{e}_s$) with 15%, *semantics2vec* (e_s) with 11%, and *math tf-idf* ($\hat{e}_m$) with 4%. On average, there were 3 matches per formula from 3 different documents. Overall, for $34/50 = 68\%$ of the sample formulae, we could retrieve equivalent representations. Finally, we listed the five top name candidates from the surrounding text and evaluated whether they contain a suitable name for the Formula Concept to be seeded as a Wikidata item. For our 50 examples, we achieve a recall of $36/50 = 72\%$ for the formula name. Furthermore, for $41/50 = 82\%$ of the retrieved name candidates, there was a Wikidata QID available to tag the formula concept.

6 Future Work

Having launched FCD as a method for tagging formulae with Wikidata QIDs, we can now employ FCR to identify formulae within STEM documents using their constituting parts (operators and identifiers) in a SPARQL query[10]. However, since at the moment only less than 4000 formulae are seeded into Wikidata [19] and storing multiple representations as "defining formula" of the same formula concept item is not endorsed, we argue for the creation of a specific Wikidata-attached *Formula Concept Database*. It should include formalized *augmentation* to generate equivalent forms using, e.g., commutations, additional sub- and superscripts, unit and reference frame variations, etc. Most importantly, a method for inferring substitutions or implicit terms needs to be developed.

Hubble's law (Q179916)	equation of state (Q214967)
$p = \omega\rho$	$\dot{a} = aH$
$p = \kappa\rho$	$H_i = \dot{R}/R$
$\omega = p/\rho$	$H = \dot{a}/a$
$p_d = \omega\rho_d$	$H(t) = \dot{a}/a$

Fig. 2: Clustering equivalent representations of formulae in the semantic space as named Formula Concept Wikidata items.

This work was supported by the German Research Foundation (DFG grant GI-1259-1).

[10] W3C Recommendation: `https://www.w3.org/TR/rdf-sparql-query`

Table 1: Formula Concept Discovery (FCD). Top-50 results of a cross-document duplicate search in the subject class astro-ph of the NTCIR arXiv dataset. Equivalent formulae are retrieved to bundle concept candidates using a k-nearest neighbor (kNN) recommendation, while comparing the relative success s of different encodings (*math2vec*: e_m, *math tf-idf*: $\hat{e}_m$, *semantics2vec*: e_s, *semantics tf-idf*: $\hat{e}_s$). The number of duplicates d and originating distinct documents $\hat{d}$ are shown as well as a retrieved sample formula. Furthermore, it is evaluated whether the first five words of the surrounding text are candidates for the name of the formula and whether a Wikidata QID is available.

#	Formula	Name (QID)	d / $\hat{d}$	$s_{e_m}, s_{\hat{e}_m}, s_{e_s}, s_{\hat{e}_s}$	Encoding: sample formula	Name candidates from surrounding text
1	$H = \dot{a}/a$	hubble parameter (Q179916)	32 / 32	0.0, 0.1, 0.0, 0.9	$\hat{e}_s$: $H_i = \dot{R}/R$	hubble, parameter, time, factor, equations
2	$p = \omega\rho$	equation of state (Q214967)	6 / 5	0.3, 0.0, 0.1, 0.6	e_s: $p_d = w\rho_d$	equation, state, quintessence, expansion, pressure
3	$\omega = p/\rho$	accelerating universe (Q1049613)	4 / 3	0.7, 0.0, 0.0, 0.3	e_m: $p = \omega\rho$	universe, accelerating, indefinitely, strain, values
4	$p = -A/\rho^\alpha$	dark fluid (Q5223514)	4 / 4	0.7, 0.0, 0.3, 0.0	e_m: $p = -\frac{A}{\rho^\alpha}$	chaplygin, gas, dark, generalized, fluid
5	$p_d = w\rho_d$	dark energy (Q18343)	4 / 3	0.3, 0.0, 0.3, 0.3	e_s: $p_X = \omega_X \rho_X$	energy, dark, equation, represent, pressure
6	$H = \dot{a}/a$	N/A (Q179916)	4 / 4	0.4, 0.1, 0.2, 0.3	$\hat{e}_m$: $\mathcal{H} = a'/a$	scale, factor, usual, equation, state
7	$k = \lvert\mathbf{k}\rvert$	wavenumber (Q192510)	3 / 3	0.8, 0.0, 0.2, 0.0	e_m: $k = \lvert\boldsymbol{k}\rvert$	oscillatory, behavior, depend, time, wavenumber
8	$f = e^{-\phi}R$	N/A (N/A)	3 / 2	1.0, 0.0, 0.0, 0.0	e_m: $f(\phi) = e^{-\phi}R$	string, lowenergy, effective, action, theory
9	$p = \kappa\rho$	equation of state (Q214967)	3 / 2	0.3, 0.0, 0.7, 0.0	e_s: $p_D = w(z)\rho_D$	equation, state, ary, patch, exceeds
10	$w = p_X/\rho_X$	equation of state (Q214967)	3 / 3	0.6, 0.0, 0.1, 0.3	e_m: $p_X = w_X \rho_X$	equation, state, dark, energy, wmap
11	$\mu = m_p/m_e$	proton-to-electron mass ratio (Q2912520)	3 / 3	1.0, 0.0, 0.0, 0.0	e_m: $m_i = \mu m_p$	ratio, proton, electron, masses, technique
12	$\phi_c = M/g$	critical value (Q2189464)	3 / 3	0.0, 0.0, 0.0, 0.0	N/A	field, critical, value, takes
13	$p = -\frac{A}{\rho^\alpha}$	chaplygin gas (Q5073250)	3 / 3	0.8, 0.0, 0.0, 0.2	e_m: $p = -A\rho^{-\alpha}$	state, generalized, chaplygin, gas, equation
14	$p = \alpha\rho$	polytropic gas (Q831024)	3 / 2	0.7, 0.0, 0.2, 0.2	$\hat{e}_s$: $w_\alpha = p_\alpha/\rho_\alpha$	constant, gas, cosmological, matter, polytropic
15	$M = \widetilde{M}/\varGamma$	connected manifold (Q2721559)	3 / 3	0.0, 0.0, 0.0, 0.0	N/A	multiply, connected, equally, quotient, manifolds
16	$g(a) = \triangle(a)/a$	dark energy (Q18343)	3 / 2	1.0, 0.0, 0.0, 0.0	e_m: $g(a) = \varDelta(a)/a$	models, dark, energy, growth, history
17	$\alpha = dn_s/d\ln k$	N/A (Q192510)	3 / 3	1.0, 0.0, 0.0, 0.0	e_m: $dn_s/d\ln k = \alpha_s$	introduced, customary, notation, comoving, wavenumber
18	$\psi = -i\theta$	N/A (N/A)	3 / 2	0.0, 0.0, 0.0, 0.0	N/A	real imaginary universe
19	$dt = a(\eta)d\eta$	N/A (Q11471)	2 / 2	0.5, 0.0, 0.3, 0.3	$\hat{e}_s$: $t = \int a(\eta)d\eta$	time, related, cosmic, relation, overdot
20	$\varDelta x_{min} = \sqrt{\beta}$	lower bound (Q21067468)	2 / 2	1.0, 0.0, 0.0, 0.0	e_m: $\varDelta x_{\min} = \hbar\sqrt{\beta}$	positive, constant, lower, bound, implies, dimensional
21	$k^i = ap^i$	modes (N/A)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	modes, comoving, obtained, scaling, coincide
22	$\varphi = \delta A_\mu$	perturbations (Q911364)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	note, valid, perturbations, gauge, theories
23	$h_{ab} = g_{ab} - n_a n_b$	metric (Q865746)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	bulk, scalar, curvature, induced, metric
24	$K = K_{ab}h^{ab}$	brane (Q385601)	2 / 2	1.0, 0.0, 0.0, 0.0	e_m: $K = K_{\alpha\beta}h^{\alpha\beta}$	vector, field, unit, normal, brane
25	$v = \sqrt{\lvert dp/d\rho\rvert}$	equation of state (Q214967)	2 / 2	1.0, 0.0, 0.0, 0.0	e_m: $v_c = \sqrt{dp_c/d\rho_c}$	equation, state, suggests, effective, velocity
26	$Q = \sqrt{G}M$	limit (Q246639)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	limit, rhoades, value, write
27	$\zeta = H\delta\phi/\dot{\phi}$	N/A (Q10886678)	2 / 2	1.0, 0.0, 0.0, 0.0	e_m: $\mathcal{R} = (H/\dot{\phi})\delta\phi_\psi$	curvature, perturbation, uniform, density, valid
28	$m_\gamma = e/\sqrt{\pi}$	photon mass (Q3198)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	photon, mass, gauge, mechanism, schwinger
29	$d\eta = dt/a(t)$	conformal time (Q2482717)	2 / 2	0.6, 0.0, 0.1, 0.3	$\hat{e}_s$: $t = \int a(\eta)d\eta$	conformal, time, ase, figure, fig
30	$T_g = H_o t_g$	N/A (Q126818)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	dimensionless, factor, eq, extragalactic, object
31	$\mathcal{H} = a'/a$	N/A (Q179916)	2 / 2	0.7, 0.0, 0.1, 0.2	$\hat{e}_s$: $H = \dot{a}/a$	conformal, time, background, scale, factor
32	$\theta = A\exp(-\zeta t)$	exponential decrease (Q574576)	2 / 2	0.0, 1.0, 0.0, 0.0	$\hat{e}_m$: $\psi(t,r) = \psi(r)\exp(-i\omega t)$	decreases, exponentially, slowly
33	$p_i = \omega_i \rho_i$	N/A (N/A)	2 / 2	0.7, 0.0, 0.1, 0.1	e_s: $w_{\mathrm{X}} = p_{\mathrm{X}}/\rho_{\mathrm{X}}$	case, expected, current, observations, restrict
34	$i\partial_t \varPhi = H\varPhi$	schrödinger evolution (Q165498)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	evolution, shrödinger
35	$H(t) = \dot{a}/a$	N/A (Q179916)	2 / 2	0.8, 0.1, 0.0, 0.1	e_m: $\dot{a} = aH$	data, scale, function, combined, sn
36	$p_\varLambda = -\rho_\varLambda$	dark energy (Q18343)	2 / 2	1.0, 0.0, 0.0, 0.0	e_m: $p_D = -\rho_D$	dark, contributions, matter, energy, matterdominated
37	$P_M = w\rho_M$	equation of state (Q214967)	2 / 2	0.6, 0.0, 0.3, 0.1	e_s: $p_x = w\rho_x$	pressure, write, related, equation, state
38	$f_\nu = \rho_\nu/\rho_d$	neutrino (Q2126)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	matter, neutrino
39	$A_t = rA_s$	fluctuation (Q5462624)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	fluctuation
40	$p_m = \gamma\rho_m$	nonrelativistic matter (Q55921784)	2 / 2	1.0, 0.0, 0.0, 0.0	e_m: $\gamma = p/\rho$	matter, components, universe, nonrelativistic, ordinary
41	$\varOmega_i = \rho_i/\rho_c$	expansion rate (N/A)	2 / 2	1.0, 0.0, 0.0, 0.0	e_m: $\varOmega = \rho/\rho_{\mathrm{crit}}$	universe, constant, rate, expansion, variables
42	$P(k) = Ak^n$	inflation (Q273508)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	fluctuations, field, inflation, universe, inflationary
43	$L_I = M(\tau)\phi[x(\tau)]$	N/A (N/A)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	idea, quantitative, viewpoint, arises, study
44	$L = \kappa h_{ab}T^{ab}$	N/A (N/A)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	standard coupling
45	$w_i = P_i/\rho_i$	equation of state (Q214967)	2 / 2	0.7, 0.0, 0.2, 0.1	$\hat{e}_s$: $w_\alpha = p_\alpha/\rho_\alpha$	relative, contributions, components, equations, state
46	$\bar{M} = B/C$	N/A (N/A)	2 / 2	0.3, 0.0, 0.3, 0.3	e_s: $\bar{M} = \frac{B}{C}$	minimum
47	$\varPsi = \varPsi_\ell + \varPsi_s$	N/A (N/A)	2 / 2	0.0, 0.0, 0.0, 0.0	N/A	split, dropped, note, long, short
48	$z = a\dot{\phi}/H$	equation (Q21086835)	2 / 2	0.7, 0.0, 0.0, 0.3	$\hat{e}_s$: $z_q = a\dot{\phi}/H$	quantity, equation
49	$u^\mu = dx^\mu/d\tau$	comoving fluid (Q5462744)	2 / 2	1.0, 0.0, 0.0, 0.0	e_m: $k^\mu = dx^\mu/dv$	cosmological, fundamental, observer, comoving, fluid
50	$\dot{\phi} = -W_\phi$	firstorder differential equation (Q11214)	2 / 2	1.0, 0.0, 0.0, 0.0	e_m: $\dot{\chi} = -W_\chi$	equation, firstorder, differential, scale, factor

References

1. Aizawa, A., Kohlhase, M., Ounis, I., Schubotz, M.: NTCIR-11 math-2 task overview. In: NTCIR. National Institute of Informatics (NII) (2014)
2. Arbab, A.I.: Derivation of dirac, klein-gordon, schrödinger, diffusion and quantum heat transport equations from a universal quantum wave equation. EPL (Europhysics Letters) **92**(4), 40001 (2010)
3. Cohl, H.S., McClain, M.A., Saunders, B.V., Schubotz, M., Williams, J.C.: Digital repository of mathematical formulae. In: CICM. Lecture Notes in Computer Science, vol. 8543, pp. 419–422. Springer (2014)
4. Detweiler, S.: Klein-gordon equation and rotating black holes. Physical Review D **22**(10), 2323 (1980)
5. Guidi, F., Coen, C.S.: A survey on retrieval of mathematical knowledge. Mathematics in Computer Science **10**(4), 409–427 (2016)
6. Haroun, K.M., Yagob, A.A.M., Allah, M.D.A.: Derivation of klein–gordon equation for frictional medium. American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) **38**(1), 1–6 (2017)
7. Kaloyerou, P., Vigier, J.: Evolution time klein-gordon equation and derivation of its nonlinear counterpart. Journal of Physics A: Mathematical and General **22**(6), 663 (1989)
8. Le, Q.V., Mikolov, T.: Distributed representations of sentences and documents. In: ICML. JMLR Workshop and Conference Proceedings, vol. 32, pp. 1188–1196. JMLR.org (2014)
9. Lozier, D.W.: NIST digital library of mathematical functions. Ann. Math. Artif. Intell. **38**(1-3), 105–119 (2003)
10. McKiernan, G.: arxiv.org: the los alamos national laboratory e-print server. International Journal on Grey Literature **1**(3), 127–138 (2000)
11. Morawetz, C.S.: Time decay for the nonlinear klein-gordon equation. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **306**(1486), 291–296 (1968)
12. Pecher, H.: Nonlinear small data scattering for the wave and klein-gordon equation. Mathematische Zeitschrift **185**(2), 261–270 (1984)
13. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
14. Rehurek, R.: Scalability of Semantic Analysis in Natural Language Processing. Ph.D. thesis, Masarykova univerzita, Fakulta informatiky (2011)
15. Scharpf, P., Schubotz, M., Gipp, B.: Representing mathematical formulae in content mathml using wikidata. In: BIRNDL@ SIGIR. pp. 46–59 (2018)
16. Scharpf, P., Schubotz, M., Gipp, B.: Representing mathematical formulae in content mathml using wikidata. In: BIRNDL@SIGIR. CEUR Workshop Proceedings, vol. 2132, pp. 46–59. CEUR-WS.org (2018)
17. Schubotz, M., Greiner-Petter, A., Scharpf, P., Meuschke, N., Cohl, H.S., Gipp, B.: Improving the representation and conversion of mathematical formulae by considering their textual context. In: JCDL. pp. 233–242. ACM (2018)
18. Schubotz, M., Grigorev, A., Leich, M., Cohl, H.S., Meuschke, N., Gipp, B., Youssef, A.S., Markl, V.: Semantification of identifiers in mathematics for better math information retrieval. In: SIGIR. pp. 135–144. ACM (2016)

19. Schubotz, M., Scharpf, P., Dudhat, K., Nagar, Y., Hamborg, F., Gipp, B.: Introducing mathqa-a math-aware question answering system. In: Proceedings of the ACM/IEEECS Joint Conference on Digital Libraries (JCDL), Workshop on Knowledge Discovery, Fort Worth, USA (2018)
20. Strauss, W., Vazquez, L.: Numerical solution of a nonlinear klein-gordon equation. Journal of Computational Physics **28**(2), 271–278 (1978)
21. Tiwari, S.: Derivation of the hamiltonian form of the klein-gordon equation from schrödinger-furth quantum diffusion theory: Comments. Physics Letters A **133**(6), 279–282 (1988)
22. Tretyakov, O.A., Akgun, O.: Derivation of klein-gordon equation from maxwell's equations and study of relativistic time-domain waveguide modes. Progress In Electromagnetics Research **105**, 171–191 (2010)
23. Vrandecic, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Commun. ACM **57**(10), 78–85 (2014)
24. Youssef, A., Miller, B.R.: Deep learning for math knowledge processing. In: Rabe, F., Farmer, W.M., Passmore, G.O., Youssef, A. (eds.) Intelligent Computer Mathematics - 11th International Conference, CICM 2018, Hagenberg, Austria, August 13-17, 2018, Proceedings. Lecture Notes in Computer Science, vol. 11006, pp. 271–286. Springer (2018). https://doi.org/10.1007/978-3-319-96812-4_23, `https://doi.org/10.1007/978-3-319-96812-4_23`
25. Yucong, D., Cruz, C.: Formalizing semantic of natural language through conceptualization from existence. International Journal of Innovation, Management and Technology **2**(1), pages–37 (2011)