A Semantic Calculus: Common Sense Reasoning for Information Systems

Rory Lewis
Computer Science Department
University of Colorado Colorado Springs
rlewis5@uccs.edu

ABSTRACT
With the continuing goal of discovering a theoretical foundation to have information systems think more like humans, this paper shows that as information systems explode in size and complexity, the notion that technologies such as Big Data, Deep Neural Networks, Graphic Processing Units, Artificial Intelligence and Machine Learning can keep up to the pace that information systems are exploding, has failed. 80% of the trillions of petabytes being stored on information systems every day, are being stored by unsophisticated end users, and that's just the start of the chaos. In essence we are drowning in data and we need a new technology to take us the surface so we can breathe. In this paper, the author explores the viability of combining sentiment analysis, and semantic intelligence into the paradigm of a common-sense semantic calculus, specifically designed for Information Systems.

CCS Concepts

Theory of computation →Semantics & reasoning
Mathematics of computing → Information theory

Keywords
information systems management; machine intelligence; artificial intelligence; cognitive intelligence; artificial consciousness

1. INTRODUCTION
We define an Information system as an integrated system of components for collecting, storing, and processing data for the goal of procuring knowledge [1]. Information systems process the world's financial systems, manage data for Amazon Web Services (AWS), Google and Alibaba to name a few. In fact, information systems are called the *new economy* because the explosion of information and big data is now driving the biggest economies [2], [3]. As these information systems 1) become increasingly complex [4] and 2) the volume of data being generated and stored grows exponentially, the technology to manage these exploding systems is not keeping up [5], [6]. Müller *et al*., provide a convincing argument that integrating big data analytics (BDA) with these new, massive, chaotic, complex information systems, has not lived up to the hype. Furthermore, the Information System Analysists (ISAs), felt that BDA's tools such as predictive modelling, natural language processing were too obtuse and difficult for them to understand [6]. Similarly, artificial intelligence (AI) is also not being welcomed with open arms by the ISAs they feel that the AI techniques focus on prediction instead of explanation [7], and hide how they work thus making it difficult to interpret their outputs or control [6].

ICIME 2019, September 19–21, 2019, London, UK

ACM ISBN 978-1-4503-5337-3/17/10…$15.00
DOI: https://doi.org/10.1145/3149572.3149592

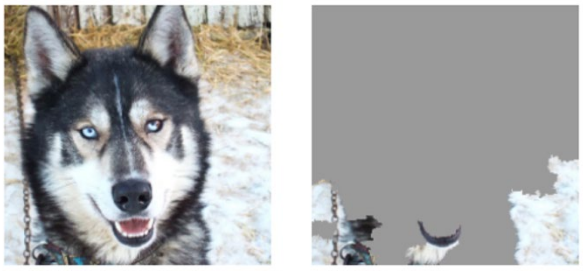
If I were to ask you, hypothetically, in a world without the internet: "*Tell me about Oliver Cromwell*", would you start looking through the records of Ghanaians? Would you go to your local library and search the obituaries in The New York Times over the last three years? No of course not. You probably know that Cromwell ruled the UK in the 1650s, with about the same power of a monarch, but he had no crown. Citing *Davis & Marcus*, [8], If I were to ask an information system, "*Who is taller, Prince William or his baby son Prince George*?" The information System would fail. As a human, it seems silly to us because *of course* Prince William is taller than his baby son. The key determinants to procuring an intelligent information system to think like a human are now reviewed.

<u>*Commonsense Knowledge*</u> consists of facts about the everyday world, such as "*lemons are sour*", that all humans are expected to know and is agnostic as it's not tied to one particular domain of knowledge [9]. <u>*Semantic intelligence*</u> is the ability to intelligently respond to a non-predetermined and non-specified ever-changing environment and have autonomous comprehension and discrimination between true and false statements. <u>*Sentiment Analysis*</u> is the computational treatment of opinions evaluations, attitudes, and emotions of humans [10] and result, for example, as an aggregation of "votes", reviews on Yelp, representation of points of disagreement and points of consensus [11]. <u>*Semantic Calculus*</u> is when synchronous grammars correctly generate logical forms. With advances in computing power, it has now become possible for intelligent systems to formalize syntax-based translation models between natural languages which has resulted in a renewed interest in synchronous grammars [12], [13].

Paper Layout: <u>*First*</u> we discuss a little deeper as to why Deep Neural Networks and other new technologies have not and will not be the save-all for Information Systems and that focusing on making an information system think like a human is a viable option. <u>*Secondly,*</u> we rationalize how, in order to tie the aforementioned commonsense determinants into an information system there needs to be a mathematical basis behind the code. Accordingly, this paper will focus, for the most part, on i) Fontaine *et al.'s* recent developments in model-theoretic tradition using modal μ-calculus [14], ii) combinatory categorial grammar (CCG) coupled with logical forms in λ-calculus [15], iii) *Cox's* theorem [16], and iv) *de Finetti's* theorem to force autonomous reasoning even when based on uncertain information [17].

2. DNNS & INFORMATION SYSTEMS

Although there has, and are, numerous efforts to integrate deep neural networks (DNNs) with Information Systems, [18], [19] their Achilles heel is that they rely on supervised learning which of course obliterates any chance on applying such a system on sparsely or erroneously labeled data. With 80% most of Information System's data being derived from users on social media [20], most of the data is therefore incomplete and erroneously labeled, which therefore wipes out the possibility of DNNs being a viable option for future Information Systems. Ribeiro *et al*., revealed an additional, well publicized weakness of



(a) Husky classified as wolf (b) Explanation

Figure 1. *Ribeiro et al.'s* Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task **[22].**

DNNs where the DNNs miraculously was able to discern wolves from huskies, a feat many humans have difficulty doing. As illustrated in **Figure 1**, it was later revealed that Google's pre-trained Inception neural network [21] happened to have most wolf images in the snow. Without human's knowing how the system trained itself intelligently, it was later discovered that the DNN had trained on a terrible classifier – the snow in the background of the wolf images. Therefore, a squirrel in the snow was also deemed by the DNN to be a wolf [22]!

3. A NEW MEANS FOR INFORMATION SYSTEMS

With AI being an unknown that only predicts the future but not the intelligence in an Information System, and with DNNs unable to render sparsely or erroneously labeled data without supervised learning, there clearly is a need to think completely differently on how we will control and mine Information Systems for knowledge in the future. This leads us to the idea that we need to make information systems not plough through trillions of petabytes by brute force to find an answer, but rather to have information systems intuitively know that *lemons are sour*, or that *Oliver Cromwell will not be in Ghanaian data base of living humans*, or that of course, *Prince William is taller than his baby son Prince George*?" In essence, we need to make information systems think like us human being do. The question is – *how*? Understanding the semantics, inuendo and nuances of language is not trivial matter for a machine. Consider Wingard's question: *The trophy doesn't fit in the suitcase because it is too big*. The question is What is too big? Answer 0: *the trophy*. Answer 1: *the suitcase.* For humans this is silly, of course it's *the trophy* but the question is how do we get a machine to know it's the trophy and not the suitcase [20]?

3.1 CCG Formalism

<u>**Hypothesis:** For a machine to think like a human, it must first be able to comprehend language like a human.</u> In other words, it needs

Category	Traditional Name	Expression
NP\S	Intransitive Verb	John Walks
NP/N	Determiner	The dog
(NP\S)/NP	Transitive Verb	John loves Mary
N/N (=A)	Adjective	Big dog

Table 1. Combinatory Categorical Grammar (CCG) CCGR: List of categories using the three primitive categories, *N*, *NP*, & *S*. For complete list *Futo* here **26.**

to semantically parse the grammar of language. One approach is that of Combinatory Categorial Grammar (CCG) [15], used in linguistics long before we tried to code machines for semantic parsing [23]. Verbs are converted into functions through connecting them with these lexical categories. For example, a transitive verb is a function from (object) NPs into predicates—that is, into functions from (subject) NPs into S [24]. Furthermore, CCG is conducive for computer algorithms because is uses logical forms of λ-calculus, see examples of CCG derivations [25] in **Table 1**. With CCG, any primitive category is a category. Therefore, if *A* and *B* are categories, then *A/B* and *B\A* are also categories [26]. Even when only employing three primitive categories *N*, *NP* and *S*, the syntactic information in **Equation 1** is transferred to a lexical entry as seen in **Equation 2** where the syntactic category makes *TV* a function [26], [27].

$$\begin{aligned} S &\rightarrow NP\ VP \\ VP &\rightarrow TV\ NP \\ TV &\rightarrow \{proved, finds, \ldots\} \end{aligned} \qquad \mathbf{1}$$

$$proved := (S\backslash NP)/NP \qquad \mathbf{2}$$

CCG is parsable and linguistically robust in terms of its grammatical formalism. It has a predicate-argument structure, a constituency-based structure rather than a dependency-based and turns out to be well suited to integrate future research for infusing a sematic calculus into future information Systems [15]. CCG defines a means to process a plurality of linguistic combinators into a sequence of lexical items. It rotates, each discretized series of combinators until none are left in the proof. Specifically, CCG is comprised of 425 lexical tags connected too a small set of 11 binary rules [28] and this is perfect for classification rules necessary for artificial intelligence. Consider the following CCG by *Steedman* that compensates for constituents of like types to conjoin to yield a single constituent of the same type [24].

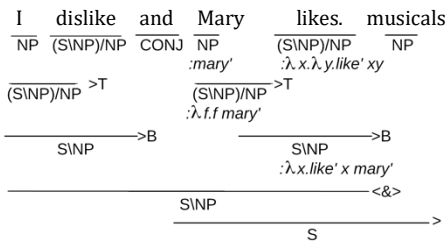
$$\begin{aligned} &Coordination: (<\&>) \\ &X\ conj \Longrightarrow X \end{aligned} \qquad \mathbf{3}$$

$$\begin{array}{ccccccc} \underline{\text{I}} & \underline{\text{loathe}} & \underline{\text{and}} & \underline{\text{detest}} & \underline{\text{opera}} & & \mathbf{4} \\ \text{NP} & \text{(S\NP)/NP} & \text{CONJ} & \text{(S\NP)/NP} & \text{NP} & & \\ & \multicolumn{3}{c}{\overline{\qquad\qquad\qquad}^{<\&>}} & & & \\ & \multicolumn{4}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad}^{>}} & \\ & & \multicolumn{3}{c}{\text{S\NP}} & & \\ \multicolumn{5}{c}{\overline{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}^{<}} & & \\ \multicolumn{5}{c}{\text{S}} & & \end{array}$$

Steedman's combinatory grammars also include type-raising rules, which turn arguments into functions over functions-over-such-arguments. Comparing **Equation 3** & **4's** Coordination example to **Equations 5** & **6's** example of Subject Type-Raising that allows the conjuncts to form which illustrates its "*order-preserving*" property that converts the NP into a rightward looking function that preserves the linear order of both the subjects and predicates [27]. This all follows Steedman's <u>Principle of Combinatory Transparency</u> which is the semantic interpretation of the category resulting from a combinatory rule is uniquely determined by the

interpretation of the slash in a category as a mapping between two sets [24], [27].

$$Subject\ Type-raising: (>T) \qquad \textbf{5}$$
$$NP : a \Rightarrow T\ (T\backslash NP)\ :\ \lambda x.\lambda y.like^{\wedge\prime}\ xy$$



6

3.2 Modal μ - Calculus

Bearing the semantic, lexical and syntactic properties of CCGs focus on the predicate logic necessary to accomplish our hypothesis. Connecting a stronger and more advanced formal logic to CCGs is possible because the lexicon of CCGs, as mentioned, allows for predicate logic. With this in mind we examine connecting a modal μ-calculus to the CCG logic component of CCGs as indicated in **Figure 2**.

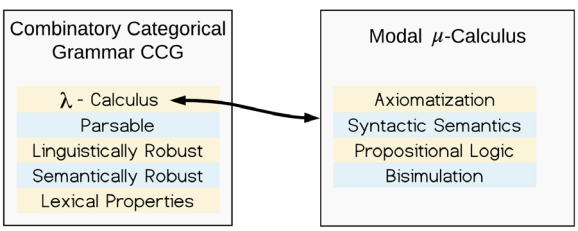


Figure 2. Combining the $\boldsymbol{\lambda}$-Calculus component of CCG formalism to the Modal μ-Calculus ecosystem.

In recent months Gaëlle Fontaine and Yde Venema at the Institute for Logic, Language & Computation at the Universiteit van Amsterdam published a novel paper that based upon the theoretical giants including but not limited to the *Los-Tarski Theorem*, the *Lyndon Theorem* and *Kozen's* work on fixpoint operators, procured an automata-theoretic model theory for modal μ-Calculus [14]. This paper utilizes a specific section of the Fontaine *et al.'s* theory and connects it to both 1) the λ-calculus component of CCG, and 2) the author's own work to create a pathway for commonsense reasoning for information systems.

Modal μ-calculus, $\mu\mathbb{ML}$, modal logic is expressively complete for monadic second-order properties that are bisimulation invariant, [29], [14]. In other words, it can sustain a connectivity to a binary relation between state transition systems that are morphing in their disposition. Specifically interesting is their introduction to a novel bipartite modal automaton $\mathbb{A}$ = (A,B,Θ; Ω) that belongs to the class $Aut\frac{W}{p}$ of finite-width automata if the one-step language associated with B is the language $1\mathbb{ML}(X\{p\}, B)$, and the one-step language associated with A is given by the following grammar:

$$\alpha ::= \mathrm{p} \mid \diamond_{\pi_0} \mid \beta \mid \alpha \wedge \alpha \mid \top \mid \alpha \vee \alpha \mid \bot \qquad \textbf{7}$$

Where ⊤ defines the proposition that $\alpha \wedge \alpha$ is universally true, ⊥ defines the proposition that $\alpha \vee \alpha$ is contradictory, π_0 Latt(A) and $\beta \in 1\mathbb{ML}$ (X \ {p},B). In words, an initialized modal automaton $\mathbb{A} = \langle a_{\mathrm{I}} \rangle$, with $\mathbb{A}$ = (A,Θ; Ω), belongs to the class $Aut\frac{W}{p}$ if A can be partitioned as $\mathbb{A} = A_0 \uplus A_1$ such that (0) a_{I} belongs to , (1) p occurs only positively in $\Theta(a)$, for $a \in A_0$, (2) p does not occur in any in $\Theta(a)$, for $a \in A_1$, (3) if $a, b \in A_0$ then a may only occur in $\Theta(b)$, for in the scope of a diamond (not a box) of modality, and (4) if $a \in A_0$ and $b \in A_1$ then a may not occur in $\Theta(b)$ [14].

Connecting the aforementioned modal μ-calculus, $\mu\mathbb{ML}$, modal logic to CCG's λ-calculus component as illustrated in **Figure 2** we turn to a variant Lewis *et al.'s* supertagging based on their stacked bi-directional long short term memory (LSTM) CCG parsing model [28] where we recall that a supertag-factored model comprised of a and b are computed so that $\mathfrak{a}(y_k)$ is the score for the word k having tag y_k.

$$\mathfrak{a}(y_{i,j}) = \sum_{k=i}^{j} \mathfrak{a}(y_k) \qquad \textbf{8}$$

$$b(y_{i,j}) = \sum_{k=1}^{i-1} \max_{y_k} \mathfrak{a}(y_k) + \sum_{kj+1i}^{N} \max_{y_k} \mathfrak{a}(y_k) \qquad \textbf{\& 9}$$

where **Equation 8** follows from the definition of the supertag factored model and **Equation 9** combines this definition with max score over all words an upperbound on the word score. This is illustrated, by going back and applying it to Steedman's sentence in **Equation 3** & **4** where we say: `I loathe and detest opera`. For the purpose of the formatting of this paper using 2 small columns we will shorten it to `I loathe for` as shown in **Figure 3** where we connect the modal μ-calculus, $\mu\mathbb{ML}$, modal logic with CCG using Lewis *et al.'s* supertagging.

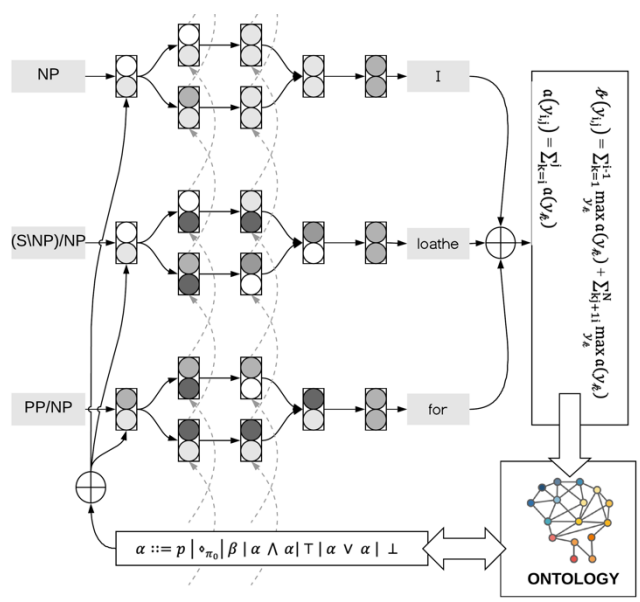


Figure 3. Connecting the modal μ-calculus, $\mu\mathbb{ML}$, modal logic with CCG using Lewis *et al.'s* supertagging and linking ontology allows sustained a connectivity to a binary relation between the state transition of `I loathe for` morphing, for example into `I` ***love*** `for` in the supertagging of **Equations 8 9** ensure connectivity.

4. CONCLUSION & FUTURE WORK

The author has structured **Figure 3** and is in the process of creating a large enough ontology for testing the system. The theoretical portion has been tested in many ways and many versions and he feels confident that after many iterations the system is working when manually super

positioning fragments of an ontology. For example on **Figure 3** the author switches I loathe for morphing, into I *love* for and the supertagging works well in conjunction with p does not occur in any in $\Theta(a)$, for $a \in A_1$, however, a larger ontology of tens of thousands of words will have to be used for the other tests.

Essentially the hypothesis seems to be correct, the commonsense will eventually build as more and kore tags are super positioned, juxtaposed, trained and learned.

5. REFERENCES

[1] E. Gupta, "Information system," *Bajaj, Ankit 197 Bakry, Mohamed Abd El Latif 28 Bala, Shashi 414 Baporikar, Neeta*, vol. 118, p. 97, 2000.

[2] L. Paganetto, *Knowledge economy, information technologies and growth*. Routledge, 2017.

[3] D. N. Chorafas, *Enterprise architecture and new generation information systems*. CRC Press, 2016.

[4] R. Kitchin, *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, 2014.

[5] D. F. Sittig and H. Singh, "A new socio-technical model for studying health information technology in complex adaptive healthcare systems," in *Cognitive informatics for biomedicine*, Springer, 2015, pp. 59–80.

[6] O. Müller, I. Junglas, J. vom Brocke, and S. Debortoli, "Utilizing big data analytics for information systems research: challenges, promises and guidelines," *European Journal of Information Systems*, vol. 25, no. 4, pp. 289–302, Jul. 2016.

[7] G. Shmueli and O. R. Koppius, "Predictive analytics in information systems research," *MIS quarterly*, pp. 553–572, 2011.

[8] E. Davis and G. Marcus, "Commonsense reasoning and commonsense knowledge in artificial intelligence," *Commun. ACM*, vol. 58, no. 9, pp. 92–103, Aug. 2015.

[9] D. Williams, "Mind as a Matter of Fact," *The Review of Metaphysics*, vol. 13, no. 2, pp. 203–225, 1959.

[10] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[11] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[12] Y. W. Wong and R. Mooney, "Learning synchronous grammars for semantic parsing with lambda calculus," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 2007, pp. 960–967.

[13] H. Ait-Kaci, "A lattice theoretic approach to computation based on a calculus of partially ordered type structures (property inheritance, semantic nets, graph unification)," 1984.

[14] G. Fontaine and Y. Venema, "Some model theory for the modal μ-calculus: syntactic characterisations of semantic properties," *Logical Methods in Computer Science ; Volume 14*, p. Issue 1; 18605974, 2018.

[15] M. Steedman, *The syntactic process*, vol. 24. MIT press Cambridge, MA, 2000.

[16] R. T. Cox, "Probability, frequency and reasonable expectation," *American journal of physics*, vol. 14, no. 1, pp. 1–13, 1946.

[17] E. Davis, "The Relevance of Proofs of the Rationality of Probability Theory to Automated Reasoning and Cognitive Models https://arxiv.org/abs/1310.1328," Oct. 2013.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[19] G. Hinton *et al.*, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[20] T. H. Trinh and Q. V. Le, "A Simple Method for Commonsense Reasoning," *arXiv:1806.02847 [cs]*, Jun. 2018.

[21] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[22] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," *arXiv:1602.04938 [cs, stat]*, Feb. 2016.

[23] L. S. Zettlemoyer and M. Collins, "Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars," *arXiv preprint arXiv:1207.1420*, 2012.

[24] M. Steedman, "A Very Short Introduction to CCG," p. 8.

[25] P. Liang, "Learning Executable Semantic Parsers for Natural Language Understanding," *arXiv:1603.06677 [cs]*, Mar. 2016.

[26] H. Futo, "Categorial Grammar with Features and the Parser on Web pages," in *Proceedings of the 14th Pacific Asia Conference on Language, Information and Computation*, Waseda University International Conference Center, Tokyo, Japan, 2000, pp. 79–86.

[27] M. Steedman and J. Baldridge, "COMBINATORY CATEGORIAL GRAMMAR," p. 62.

[28] M. Lewis, K. Lee, and L. Zettlemoyer, "Lstm ccg parsing," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 221–231.

[29] D. Janin and I. Walukiewicz, "On the expressive completeness of the propositional mu-calculus with respect to monadic second order logic," in *International Conference on Concurrency Theory*, 1996, pp. 263–277.

Your Name	**Title***	**Research Field**	**Personal website**
Rory Lewis	Associate Professor	Artificial Intelligence, Mathematical Theory of Machine Intelligence. Information Systems	https://www.rorylewis.com/