LexiClone Inc. at CLEF WiQA

Ilya Geller

igeller@lexiclone.com

Abstract. In the article I explain that my method is based upon obtaining of redundant information about the Egoism of the person seeking the information, and equally redundant information about the Egoism of the creators of the texts in which the information is being sought. Also I give some hints on ontological foundation of Egoism.

Introduction.

My premise is that information is always required Egoistically[1] by the person seeking it. In other words,
- by projecting the conviction that I share with Arthur Schopenhauer [10] – that any human activity exists only as long as and in so far as that person's Egoism exists
- onto the concrete problem of the *computerized* search for textual information

I demonstrate that the search for information always supposes the presence of redundant[2] information about the Egoism of the person seeking the information, and equally redundant information about the Egoism of the creators of the texts in which the information is being sought.

System Description.

Lexical Clones. The basis of the method of Lexical Cloning[3] which I used in the running of CLEF WiQA consists in creating
- a Lexical Clone of the person seeking the information
- and Lexical Clones of the creators of the texts in which the information is being sought,

where a Lexical Clone is the summary of the predicative definitions[4] extracted from the texts created or being used by the person for whom a given Clone is created.

I affirm that an individual's Lexical Clone adequately reflects his Egoism
- through that individual's choice of certain predicative definitions when dealing with certain specific themes
- and through the unique frequency with which these predicative definitions are repeated in a Lexical Clone.

In other words, when speaking of redundancy in the recording of Egoism in a person's Lexical Clone, I mean that such a Clone contains only information that directly characterizes a person's Egoism, without any admixtures. (By admixtures I mean, for example, the various commentaries, intersperses, explanations and clarifications unavoidably brought in by psychology).

More Details on Egoism and Lexical Clones. In 2004 in my article *LexiClone Inc. at NIST TREC* [6], I said that Egoism is a person's immanent[5] striving to include in himself (both literally and figuratively) everything that is outside himself, in order to become *better*[6]. As the most obvious practical demonstration of such Egoistical striving I pointed to the fact that predicative definitions containing the adjective/preposition "in" are the ones that occur most frequently in summaries of predicative definitions (such summaries being, essentially, the same as Lexical Clones).

A year later, in 2005, my article *Differential Linguistics at NIST TREC* [5] was published, in which I outlined how I intend to use ideas and terminology borrowed from Topology, Set Theory and Differential Analysis to give my understanding of "Egoism" an ontological[7] foundation:

---

[1] The word "Egoism" comes from Latin *ego* – "I". [1 {Egoism}]

[2] Redundancy is a concept in information theory. The presence of redundancy in the recording of statements from any source of information manifests as the possibility of recording those statements in a (on average) more concise way, using the same symbols (that is, exchanging one code for another that uses the same alphabet). [2 {Redundancy}]

[3] The method for creating a Lexical Clone is described in more detail in my articles [5, 6, 7].

[4] Any word taken separately in its normal form – where "normal form" is a combination of letters in which the word appears in print in a generally accepted dictionary of that language – is a "non-predicative definition". I have called combinations of normal forms of words "predicative definitions" if they (usually, as a rule) follow the pattern noun-verb-adjective.

[5] "Immanent" (from Latin *immanens* – abiding in something, characteristic of something) is a concept referring to some characteristic (regularity) internally present in some object, event, or process. [1 {Immanently}]

[6] '…for the end should not be just any last thing, but the best.' [Aristotle, 194$^\alpha$,30]. I explain what I mean by becoming better and the best in my article *Answering Factoid and Definition Questions: On Information for an Object* that I plan to publish later at NIST TREC 2006.

[7] Ontology is the study of being in and of itself, a division of philosophy that studies the fundamental principles of being [1 {Ontology}].

- in that article I supposed that which is in the Universe is *something* and that the Universe itself is *something*;
- I also defined Cynical Metaphysics as Differential Metaphysics: in every *something* is contained an immanent striving towards its limit[8], and for all things that limit is one and the same Nothing, as *something* existing outside time, in eternity.
- the existence of Nothing implies a radically new quality of being, which cannot be described by the same means or the same terminology as the quality of the being of *something*;
- Nothing is being at what I call by analogy the "limit point"[9]: it's obvious that there exists only one single limit point, one single undifferentiated Nothing;
- every *something* is a point of accumulation[10]: points of accumulation are many and they are different and differentiable in descriptions;

Now, in the present article, I elaborate this further by saying that Egoism is the manifestation of an immanent striving in points of accumulation to become the limit point. This means that, from now on, Egoism has an ontological justification.

The Philosophy of Differential Cynicism. Egoism is the basis of Cynicism, and Cynicism is one of the three pillars on which my Philosophy of Differential Cynicism rests – the other two being Stoicism and Cyrenaism.

I. In my article *Differential Linguistics at NIST TREC* I said that the founding fathers of Cynicism are Ecclesiastes and Jeremiah. Indeed, these Old Testament philosophers supposed
- that there exists only one limit point, and that all people are subject to an immanent striving towards it
- and that attaining this limit implies the impossibility of continuing to describe that which has become the limit by the same terms used to describe that which hasn't yet become it [4].

II. From Stoicism my Differential Cynicism has derived its Cosmology[11] and its Ethics[12]:
- the Universe in Differential Cynicism is presented as a finite manifestation existing as *something* and striving to organize its parts into an expediently placed Nothing;
- in reality, in observations and sensations, there exists only *something*, which changes in time;
- there exists an unchanging Nothing, in eternity;
- there exist two beginnings – something and Nothing;
- the development of the Universe takes place cyclically – from Nothing to Nothing;
- an individual's relation to the Universe, to everything and to other individuals (that is, an individual's Ethics) depends solely on himself, and is defined solely by himself [1 {Stoicism}, 8, 9, 10, 12]

III. Cyrenaism has influenced the formation of the Aesthetics[13] of Differential Cynicism:
- no thing is objectively just or beautiful, since that is defined by means of generally accepted concepts,
- happiness is impossible in so far as it is a limit
- and as a consequence one must independently choose the justice and beauty that is statistically most appropriate to one's pursuit of happiness, which is itself unattainable. [1 {Cyrenian school}]

But a Cynic always seeks for objectively justice and beauty as long as one is a part of the Universe – otherwise one does not exist into the Universe[14].

The Role of Differential Cynicism. The practical role of Differential Cynicism in the search for information lies in its incentive to search for that information which is Egoistically required and statistically most appropriate. Or, to use the popular terminology of those who like to write about the Web on the Web, in the *personalization* of information searches.

---

[8] Limit is a mathematical concept, indicating that a certain variable, in the observed process of its change, is approaching a constant value. [2 {Limit}] I believe that the limit is Aristotle's "the best".

[9] The limit point of a set, a point of accumulation is a point in any environs of which there is at least one point of a given set different from itself. [1 {Limit point}]

[10] A point of accumulation is a point x in a topological space $X \supset M$, of which any environs contains an infinite number of points of the set M; where the environs, neighborhoods of point x in the topological space X is the set $U \subset X$, for which x is an internal point. In other words, the neighborhoods is a set which contains an open set containing x. [2 {Point of Accumulation, Environs, Topology}]

[11] Cosmology is a domain of science which studies the Universe as a whole and cosmic systems as parts of that whole. [1 {Cosmology}]

[12] The core of Cynical Ethics lies in the Cynical appraisal of what will happen if something is included into the composition of one's self. [7]

[13] Cynical Aesthetics is the Cynical analysis of how something can be included in one's self -- the active mechanics of such a process.

[14] For we are saved by hope: but hope that is seen is not hope – for what a man seeth, why doth he yet hope for? [{Paul, 'To the Romans', 8&24}]

Berkeleyan Semantics. As is well known, Berkeley's philosophy is based on the affirmation that *esse est percipi*, that is, to be is to be perceived (or to perceive). It is also generally accepted that, taken to the limit, this Berkeleyan principle leads to solipisism, for which the thinking subject is the only indubitable reality, and everything else exists only in the individual's consciousness [1 {Berkeley, Solipsism}, 3].

   As a Cynic and an adept of Differential Philosophy I say that the single and true meaning of a word exists only for – and can only be found by – an individual who has become Nothing. But since, according to the theory of Differential Cynicism, there are no such individuals, no word, consequently, has or can have the only true meaning! To search only for the generally accepted meanings of words is, therefore, a complete waste of time. And for this reason the Berkeleyan Semantics I have elaborated is based on an awareness that the meaning of a word is a probabilistic quantity.

   Consequently, I use the Compatibility instrument [7], which is aimed at finding probabilistic information when seeking information:

$$\text{Compatibility} = \left(\frac{\text{Sum} \left(\text{Weight-of-the-same-predicative-definition-User} \quad * \quad \text{Weight-of-the-same-predicative-definition-Text} \right)}{\text{sqrt} \left(\text{Sum} \left(\text{Weight-of-each-predicative-definition-User}^2 \right) \quad * \quad \text{Sum} \left(\text{Weight-of-each-predicative-definition-Text}^2 \right) \right)} \right) * 100$$

where *weights* are the frequency with which predicative definitions are repeated in sentences, paragraphs, and texts.
Synonymic Expansion. In searching for information I elaborated the mechanism of synonymic expansion[15]: all possible synonyms and antonyms are found for every word in the question. Then all possible predicative definitions using them are put together, after which they are sifted through the Lexical Clone of one that is seeking the information.

Description of the Runs Submitted

   In the process of searching for textual information in CLEF WiQA I used, in order to create Lexical Clones of those ostensibly seeking the information, a selection of Wikipedia articles given as the answer to a question posed on CLEF WiQA. Then I operated a search by means of questions that had been synonymically expanded beforehand, together with these Lexical Clones of Wikipedia texts. But since the task was announced on May 24 and began on June 19 I was unable to put into practice a fundamental rule of Differential Linguistics – that information must be sought in the *paragraphs* of a text, not in its *sentences*. Indeed, I was just physically unable to find the time to break the Wikipedia texts down into paragraphs, and looked for the information in sentences. As a result, the search for information in CLEF WiQA by means of the Lexical Cloning method was reduced to finding sentences containing the very same words as in the question. This, of course, cannot under any circumstances be called an information search.

Results Plus Analysis of the Results.

   It is possible that, if CLEF WiQA continues and I can dispose of more time, I will be able to put my Lexical Cloning method to full use.

Conclusion.

   It is evident that, in the search for textual information, the creation of Lexical Clones is redundant. Indeed, even if one takes for a given that there exist a science capable of describing a person's Egoism, it is obvious that such a description needs to be translated into Natural Language. And such translation, as practice shows, is very ineffective.

References.
[1] Encyclopedical Dictionary of Philosophy. 1983: Soviet Encyclopedia. Moscow.
[2] Encyclopedical Dictionary of Mathematics. 1988: Soviet Encyclopedia. Moscow.
[3] Berkeley, G. 1957: *A Treatise Concerning the Principles of Human Knowledge*. The Bobbs-Merill Company Inc. London.
[4] *Bible*.
[5] Geller, I. 2005: *Differential linguistics at NIST TREC*. The Fourteenth Text REtrieval Conference Proceedings (TREC 2005). Washington.
[6] Geller, I. 2004: *LexiClone Inc. and NIST TREC.* Washington: NIST Special Publication 500-261.The Thirteenth Text REtrieval Conference Proceedings (TREC 2004).
[7] Geller, I. 2003: *The role and meaning of predicative and non-predicative definitions in the search for information*. NIST Special Publication 500-255. *The Twelfth Text REtrieval Conference (TREC 2003)*. Washington.
[8] Mates, B. 1973: *Stoic logic*. Los Angeles.
[9] Rist, John M. 1969: *Stoic Philosophy.* Cambridge University Press. Cambridge.
[10] Rist, John M. (ed.). 1978: *The Stoics.* University of California Press. Berkeley, Los Angeles, and London.
[11] Schopenhauer, A. 1965: *On The Basis of Morality.* Library of Liberal Arts, Trans. E.F.J. Payne. Indianapolis.
[12] Watson, G. 1966: *The stoic theory of knowledge*. Belfast.

---

[15] Described in more detail in my US Patent No. 199.067.