

テキスト文の自動意味抽出システム (Text Semantics Analyzer)

(株)サイバーエッジ
2009年3月16日(月)

Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

1

テキスト文の自動意味抽出システムの狙い

●OWL記述されているデータが有るならば、意味検索を行なうことが可能であり、有用であることは多くの人が認めている。



●既存の情報を如何にして効率的にOWL記述にマッピングできるかが、**課題**となっている。



●XMLデータなど構造化されたデータをOWLデータにマッピングすることは、比較的容易であり、多くのツールが開発されている。
●しかし、既存の情報の中でXML記述されている情報は、極一部で、ニュース記事や新聞記事など、圧倒的に膨大な情報がプレーンテキストの形式で発信されている。



●これ等の情報は構造化された情報では無い為、人間可読(ヒューマンリーダブル)ではあるが、コンピュータなどの機械で意味を汲取することは、色々なチャレンジが行なわれてきたが、難しかった。



●Text Semantics Analyzerは、**意味モデル**を踏まえたプレーンテキストの自動意味抽出を行なう。

Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

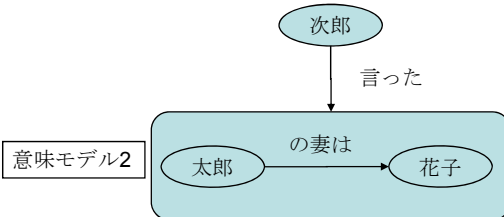
2

文の意味モデルとは？

文1 「太郎」の妻は「花子」です。



文2 「「太郎」の妻は「花子）」と「次郎」が言った。



補足説明

この様に、ノードがトリプルになることをRDFではReificationと言う。

因みに、文2は、次の如く言い換えることが可能

「次郎」は、「「太郎」の妻は「花子）」だと言った。

Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

3

理想的なオントロジの形態とは？

何故、テキスト文の意味抽出を行なうかと言うと、その意味をオントロジとして記述(OWLで記述)し、蓄積し、利用したいからである。

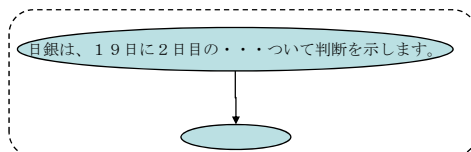
ならば、如何なる形態のオントロジが望ましいか？
文に対応したオントロジを作ろうとした時、問題になるのはオントロジの粒度である。

粒度を粗くして、句単位のオントロジを作ると、句を構成する要素の意味を汲取るのが難しくなる。(このレベルのものは、メタタグと殆ど同じとなる)
粒度を細かくして、文字単位のオントロジを作ると、文脈の意味を汲取るのが難しくなる。

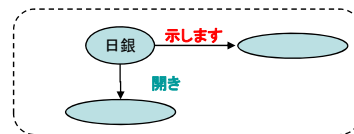
適度な粒度のオントロジが肝要である。
粒度が適切であれば、オントロジ構成要素の意味のマッピングや関連付けが容易になる。

用語及びデータ単位のオントロジが望ましい。

粒度が粗く、望ましくないオントロジ



理想的なオントロジ



Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

4

ニュース記事の意味モデル

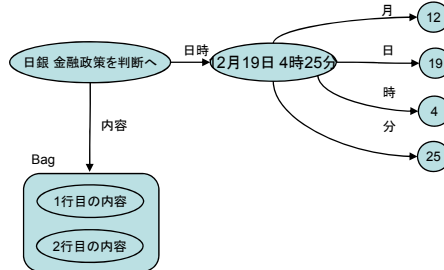
12月19日 4時25分

日銀 金融政策を判断へ

日銀は、19日に2日目の金融政策決定会合を開き、厳しさを増す景気や外国為替市場で円高ドル安が進んでいる現状を踏まえ、政策金利を引き下げる必要があるかどうかを含めた金融面からの追加の景気下支え策について判断を示します。

日銀の金融政策決定会合では、18日に、執行部から急速に厳しさを増している国内の景気の現状や、外国為替市場で円高ドル安が進んでいることなどが報告されました。2日目の19日の会合では、景気の現状について、これまでの「停滞色が強まっている」という認識から、「悪化している」に改めることで最終的に調整します。そのうえで、会合では、金融面からの追加の景気下支え策を検討します。具体的には、ことし10月に年0.3%に下げた政策金利をさらに引き下げるかどうかや、日銀が、社債の一種であるCP=コマーシャルペーパーを買ったり、金融機関から買い入れる国債を増額したりするなど、企業や金融市場に資金が行き渡るようにする対策に踏み込むかどうか議論の見通しです。今月16日には、アメリカの中央銀行に当たるFRB=連邦準備制度理事会が、事実上のゼロ金利政策に踏み切るなど、各国が思い切った政策対応に乗り出すなかで、日銀がどのような判断を示すのかが注目されます。

(注) 日本放送協会(NHK)のニュース記事を使っています。



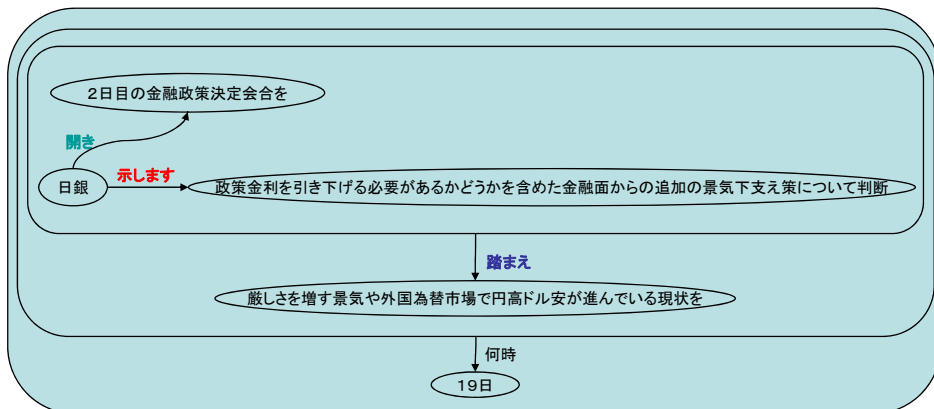
(注) この資料の意味モデルは一つの例であり、他の意味モデルも考え得る。

Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

5

ニュース記事の一行目の意味モデル

日銀は、19日に2日目の金融政策決定会合を開き、厳しさを増す景気や外国為替市場で円高ドル安が進んでいる現状を踏まえ、政策金利を引き下げる必要があるかどうかを含めた金融面からの追加の景気下支え策について判断を示します。

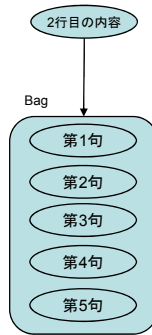


Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

6

ニュース記事の二行目の意味モデル

日銀の金融政策決定会合では、18日に、執行部から急速に厳しさを増している国内の景気の現状や、外国為替市場で円高ドル安が進んでいることなどが報告されました。2日目の19日の会合では、景気の現状について、これまでの「停滞色が強まっている」という認識から、「悪化している」に改めることで最終的に調整します。そのうえで、会合では、金融面からの追加の景気下支え策を検討します。具体的には、ことし10月に年0.3%に下げた政策金利をさらに引き下げるかどうかや、日銀が、社債の一種であるCP=商業ペーパーを買い取ったり、金融機関から買い入れる国債を増額したりするなど、企業や金融市場に資金が行き渡るようにする対策に踏み込むかどうか議論する見通しです。今月16日には、アメリカの中央銀行に当たるFRB=連邦準備制度理事会が、事実上のゼロ金利政策に踏み切るなど、各国が思い切った政策対応に乗り出すなかで、日銀がどのような判断を示すのかが注目されます。

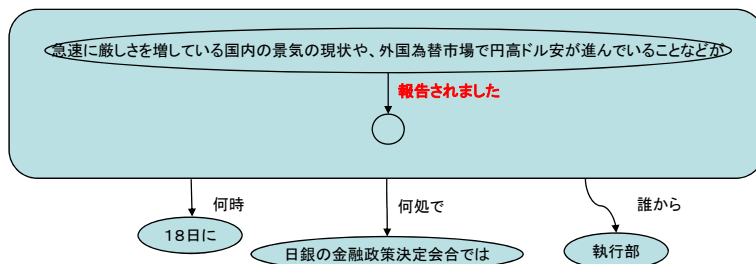


Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

7

ニュース記事の二行目の第1句の意味モデル

日銀の金融政策決定会合では、18日に、執行部から急速に厳しさを増している国内の景気の現状や、外国為替市場で円高ドル安が進んでいることなどが報告されました。

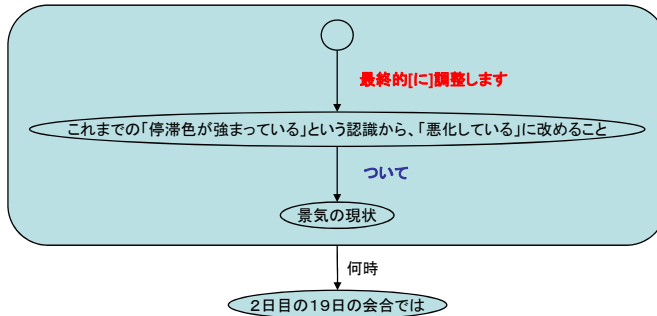


Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

8

ニュース記事の二行目の第2句の意味モデル

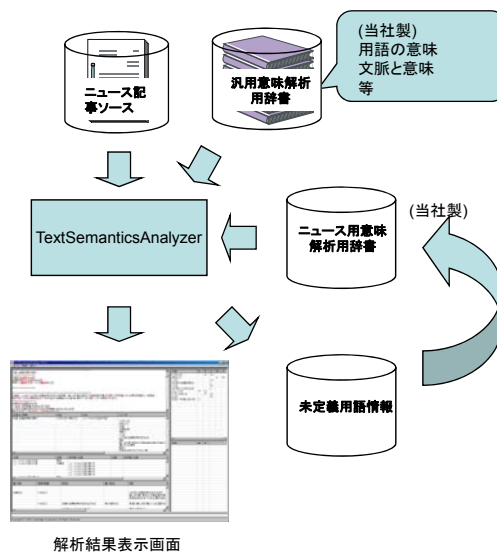
2日目の19日の会合では、景気の現状について、これまでの「停滞色が強まっている」という認識から、「悪化している」に改めることで最終的に調整します。



Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

9

テキスト文の自動意味抽出システム(Text Semantics Analyzer)の構成



Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

10

解析結果表示画面

入力テキストと用語解析結果表示ウィンドウ

基本構造とメタタグ表示ウィンドウ

句の構造表示ウィンドウ

5W1H+α情報表示ウィンドウ

時間概念用語表示ウィンドウ

計量概念用語表示ウィンドウ

Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.

Copyright(C) 2009 CyberEdge Corporation All Rights Reserved.