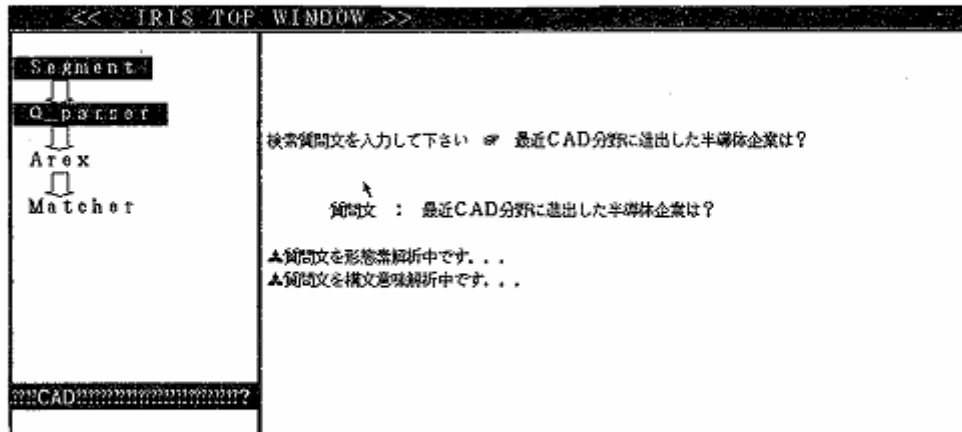


<p>題 名</p>	<p>知的情報検索実験システム (IRIS)</p>
<p>目 的</p>	<p>複数の自然言語テキスト群から、自然言語問い合わせ文の意味内容に関連するテキストを検索するシステムの試作を通じて、自然言語理解および理解内容に基づく推論技術の実証・発展を行う。</p>
<p>概要 及び 特徴</p>	<p>検索対象：情報産業界の新聞記事見出し文テキストベース</p> <p>概要：</p> <ul style="list-style-type: none"> <li>①日本語質問文から、検索されるべきテキスト文の内容を表す意味構造を生成。</li> <li>②関連命題への変換，類似概念を表す用語への展開により，質問文には現れないキーワードを含んだ検索式を生成。検索は従来型のキーワード検索システムが実行。</li> <li>③質問の内容との一致性判定のため，検索されたテキスト群の意味構造も生成。</li> <li>④質問文意味構造とテキスト文意味構造を比較・照合し，検索結果を順序付け。</li> </ul> <p>特徴：</p> <ul style="list-style-type: none"> <li>①オブジェクト指向意味表現に基づく対象分野モデル</li> <li>②対象分野のモデルを用いた構文意味解析部 <ul style="list-style-type: none"> <li>・助詞や係り受け対象の省略の多い見出し文に対処した意味関係主導型の解析。</li> </ul> </li> <li>③キーワード検索専門家の技能を用いた検索式生成エキスパートシステム <ul style="list-style-type: none"> <li>・単語間の関係だけでなく，命題間の関連性を用いて，より良い検索式を生成。</li> </ul> </li> <li>④質問文とテキスト文の一致度を判定してテキスト文を順序付けする内容照合部。 <ul style="list-style-type: none"> <li>・意味概念間の類似性を扱うことによる一致度判定。</li> </ul> </li> </ul>
<p>構 成</p>	<p>(例) CAD分野に進出した企業は？</p> <pre> graph TD     Q[質問文] --&gt; QA[質問文解析部]     QA --&gt; MS1[意味構造]     MS1 --&gt; SFG[検索式生成部]     SFG --&gt; S["(例) (進出 or 参入 or ...) and (CAD or ECAD or ...)"]     S --&gt; SKWS[従来型キーワード検索システム]     SKWS &lt;--&gt; TB[(テキストベース)]     SKWS --&gt; TA[テキスト文解析部]     TA --&gt; MS2[意味構造]     MS2 --&gt; CM[内容照合部]     MS1 --&gt; CM     CM --&gt; SR[検索結果]     SR --&gt; SR_T["(例) パソコンCAD市場, 30社参入, A社がCAD輸出, 中国に進出"]   </pre>

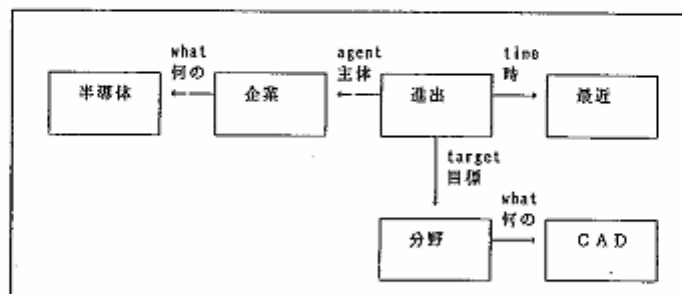
1. 質問文の入力    質問文   :   最近CAD分野に進出した半導体企業は？
2. 質問文解析     質問文解析は、形態素解析と構文意味解析からなる。



2.1 形態素解析結果

```
[word("最近", "saikin", adv, [], [], [c_time], 0, "最近", [], []), word("CAD", "cad", n, [], [[connect, left_right]], [c_form], 0, "CA
word("分野", "bunya", n, [], [[connect, left_right]], [c_field], 0, "
word("に", "ni", aux, [], [[syncat, pn], [connect, pn]], [], 0, "x_ni"
word("進出", "sinsyutsu(-suru)", vb, [], [[infl, adp], [syncat, pr
word("だ", "ta", auxvb, [], [[syncat, pran], [connect, xvcap2]], [],
word("半導体", "handoutai", n, [], [[connect, left_right]], [c_subst
word("企業", "kigyuu", n, [], [[connect, left_right]], [c_company],
word("は", "ha", aux, [], [[syncat, ptp], [connect, ptp]], [], 0, "x_h
word("?", "peried", punct, [], [], [], 0, "x_peried", [], [])
Exit
word("最近", "saikin", adv, [], [], [c_time], 0, "最近", [], [])
word("CAD", "cad", n, [], [[connect, left_right]], [c_form], 0, "CA
word("分野", "bunya", n, [], [[connect, left_right]], [c_field], 0, "
word("に", "ni", aux, [], [[syncat, pn], [connect, pn]], [], 0, "x_ni"
word("進出", "sinsyutsu(-suru)", vb, [], [[infl, adp], [syncat, pr
word("だ", "ta", auxvb, [], [[syncat, pran], [connect, xvcap2]], [],
word("半導体", "handoutai", n, [], [[connect, left_right]], [c_subst
word("企業", "kigyuu", n, [], [[connect, left_right]], [c_company],
word("は", "ha", aux, [], [[syncat, ptp], [connect, ptp]], [], 0, "x_h
word("?", "peried", punct, [], [], [], 0, "x_peried", [], [])
```

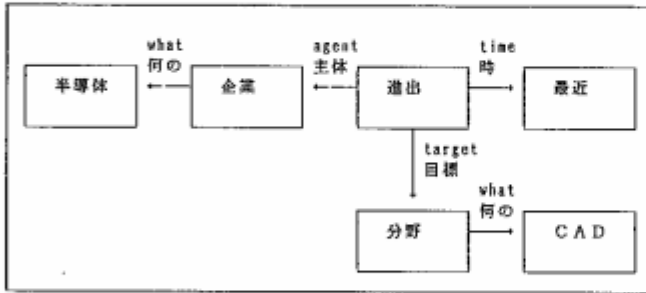
- 2.2 構文意味解析結果    質問文の解析結果は以下のような意味構造となる。



3. キーワード検索式生成

下図にあるように、質問文解析で得られた初期意味構造が、関連述語等の知識を使って変形されていき、それぞれの意味構造に対応したキーワードが生成される。これらが意味的に結合され、最終的なキーワード検索式が得られる。

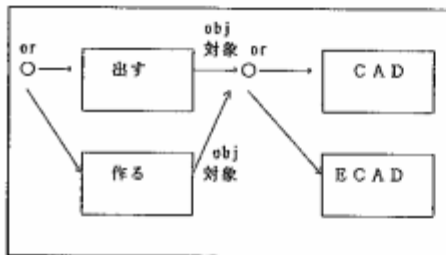
初期意味構造



キーワードへの展開

```
and [or "企業", "会社", "~ ]
     [or "半導体"]
     [or "進出", "参入", "入る", ~]
     [or "最近"]
     [or "分野", "市場"]
     [or "CAD", "CADCAM", ~ ]]
```

変形された意味構造



```
and [or [or "出す", "発表", "発売", ~]
        [or "作る", "作成", "生産", ~]]
     [or [or "CAD", "CADCAM", ~]
         [or "ECAD", ~ ] ]]
```

最終的なキーワード検索式

```
and [or "進出", "参入", "入る",
      "出す", "発表", "発売", ~ ]
     [or "CAD", "CADCAM", "ECAD", ~ ]]
```

4. キーワード検索式の実行

生成されたキーワード検索式は従来型のキーワード検索システムによって実行され、右図のような候補テキスト群が得られる。

検索結果 (CAD, CAM分野に進出した企業は?)

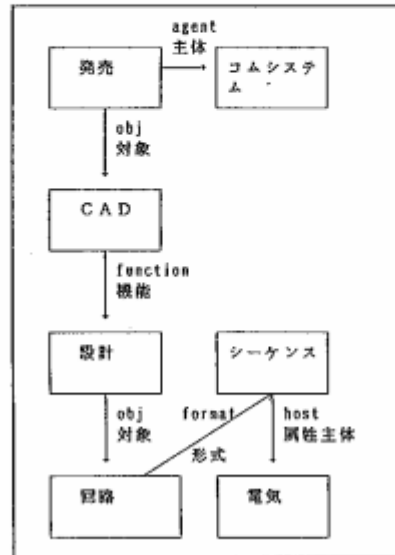
id	title
93	パソコンCAD市場◆30社以上が参入◆大型システムとも結合
94	拡大するCAE市場◆年間倍増の伸び◆主力はLSI設計分野◆メンター社など急成長を続ける◆CADに不
119	第一商会◆CAD/CAM用ソフトを販売◆英/社と合併◆社設立

5. テキスト文解析

候補テキスト群は、質問文解析とほぼ同様に文解析され、右図のような意味構造が作り出される。テキスト文解析が質問文解析と少し異なる点は、省略された助詞の補いや、見出し間に跨がった意味的關係の推定処理を行っている点である。

テキスト文 (新聞記事見出し) の意味構造

「コムシステム◆  
電気シーケンス回路設計用CADを発売」



6. 内容照合

質問文の意味構造と各候補テキストの意味構造とが比較照合され、意味的に関連が強いと判定された順に候補テキストが並べかえらるると共に、関連がないと判定されたテキストは候補から取り除かれ、これが最終的な検索結果となる。意味的関連度は、各述語の関連度や、意味的に対応する (主体、対象など) 各名詞的要素の関連度から計算される。

テキストの関連度判定

MATCHER DEBUG WINDOW			
①質問文①			
最近CAD/CAM分野に進出した半導体企業は?			
②テキスト文②			
コムシステム◆電気シーケンス回路設計用CADを発売			
質問文	テキスト文	適合率	
1	進出 VS 発売	=	257

検索結果

Match: 13 / 47 (マッチング数: 総CAD、CAM記事)			
	rate	id	title
1位	428	93	パソコンCAD市場◆30社以上が参入◆大型システムとも結合
2位	428	190	富士通◆ユーザーと共同で開発◆建築用CAD
3位	257	641	コムシステム◆電気シーケンス回路設計用CADを発売
⋮	212	684	CR Cが米社と総代理店契約◆全話型光学設計システム販売◆会話型交通騒音解析システム発売

関連度の計算方法

関連度 =

(質問文・テキスト文の述語関連度) ×

(質問文・テキスト文の対応する名詞的要素の関連度の平均)