



Semantic Search

An Active Approach to Searching

Why Semantic Search?

A better question is – why not? We are interested in what the document means, not whether it contains some key words.

There are many ways of saying the same thing, so searching using key words is unreliable – it is more reliable to search for a meaning, so the searching is not confounded by negations or multiple topics, or the inability to say exactly what you require.

We have tolerated crummy methods for a long time, because the computer resources were inadequate to the task – that has changed.

Why Not Static Search 1

Let's say you want to find something on tsunami wave theory, but not Airy waves (wind waves)

You could search for “tsunami, theory, NOT Airy”

But then you would miss

“We looked at John's work on tsunamis, but rejected it because of its use of Airy wave theory – instead we developed our own wave theory”

Static search makes several very strong assumptions – the document is short, it is only directed to the topic at hand, and it doesn't use a lot of NOTs.

In other words, it is extremely dumb. If a few hits are expected, and the person is willing to sort among them, fine. If the hits run into the tens of thousands, and it is tedious to sort among them, not fine. Worse, you have to figure out whether your search strategy will exclude the very documents you are searching for.

Why Not Static Search 2

Today it is Enron, and we want to find out if any information was omitted from reports to FERC.

We search for “report omission FERC”

A document says

“We prepared the report for FERC.

We also prepared a report to the board, leaving out some information, as it had been covered in previous board meetings.”

This seems to have the basic elements – “report”, “FERC”, and “leaving out” (a synonym for omission), but it clearly doesn’t qualify.

“We followed John’s suggestion in his email of 15th inst in preparing the report”

Someone or something has to read it (and potentially other documents) to get the sense of what it is saying – its semantics.

Freedom of Expression

Semantic search allows you to say exactly what you want – if some results come back that aren't what you want, you change how you express your wishes.

You don't need to worry about knowing the exact way something is described in the document – the meaning is searched for, not the words.

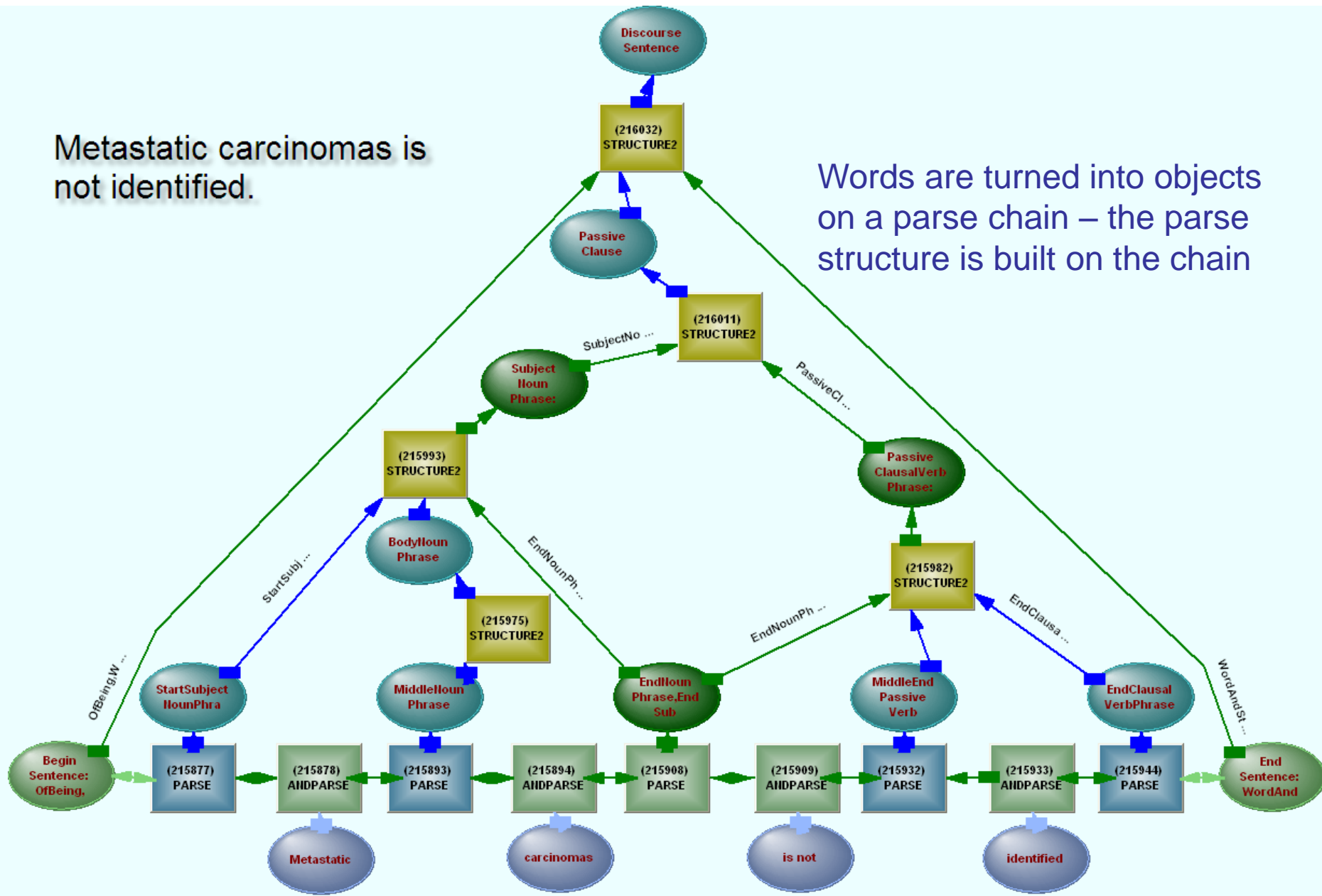
This can be done because a structure to represent the meaning of the document has already been built – the structure was built to allow any sort of search on the document because the structure is the exact structural analogue of the meaning expressed in free text.

The search structure is matched against the document structure – the matching becomes active where that is necessary – the generation of multiple possibilities and their pruning.

We Get from Words to Structure

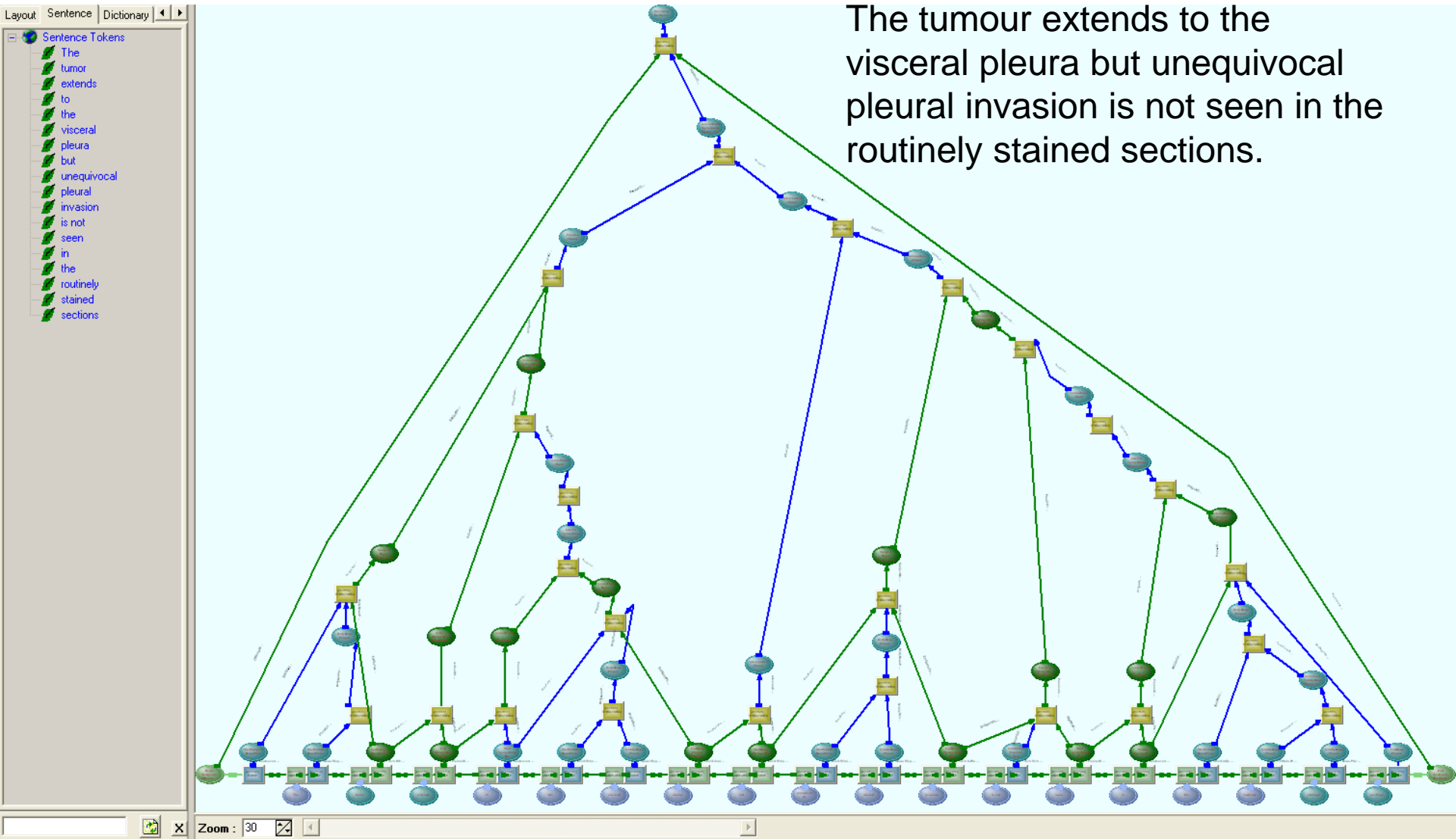
Metastatic carcinomas is not identified.

Words are turned into objects on a parse chain – the parse structure is built on the chain



More Parsing

The tumour extends to the visceral pleura but unequivocal pleural invasion is not seen in the routinely stained sections.



The tumor extends to the visceral pleura but unequivocal pleural invasion is not seen in the routinely stained sections

Definitions of Words

Main Entry: **me-tas-ta-sis** 🗣️

Pronunciation: \mə-ˈtæs-tə-səs\

Function: *noun*

Inflected Form(s): *plural* **me-tas-ta-ses** \-,sēz\ 🗣️

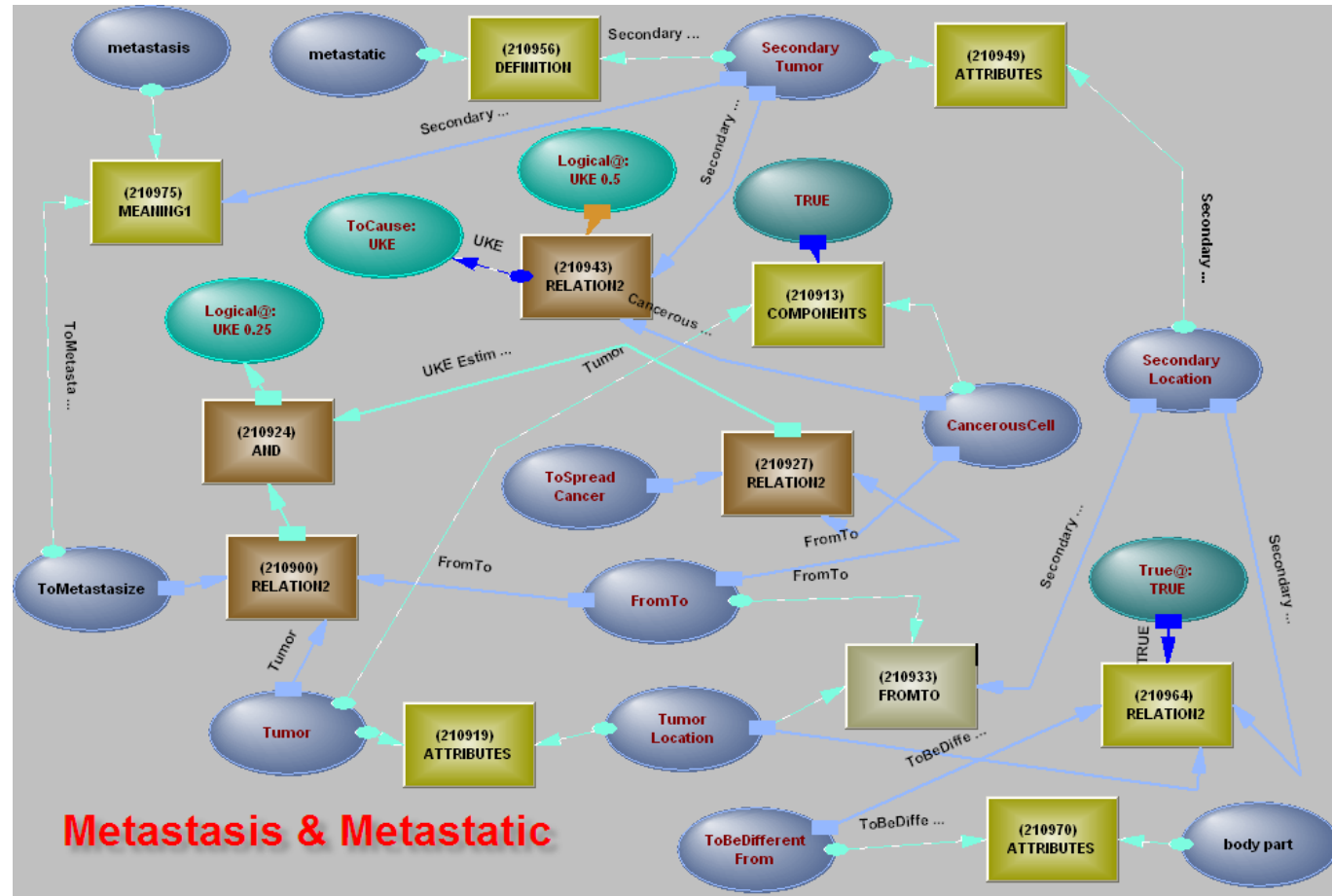
1 **a** : change of position, state, or form **b** : the spread of a disease-producing agency (as cancer cells or bacteria) from the initial or primary site of disease to another part of the body <metastases of breast cancer to bone—*Medical Physics*> ; *also* : the process by which such spreading occurs

2 : a secondary malignant tumor resulting from metastasis

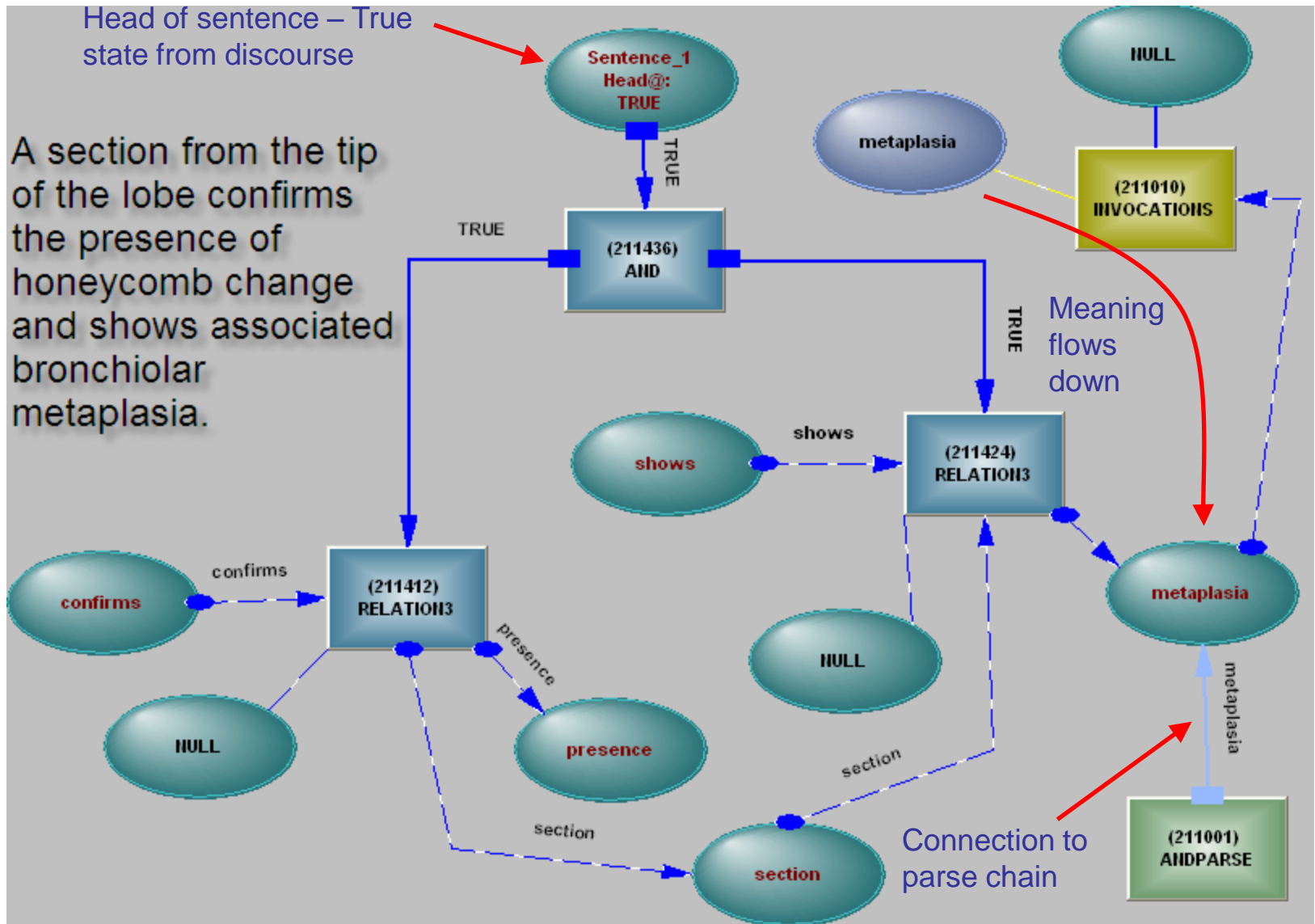
The structural definition is limited to only cancer.

There are two meanings – the process of spreading or the resultant tumor.

(the FROMTO is used for all movement relations)

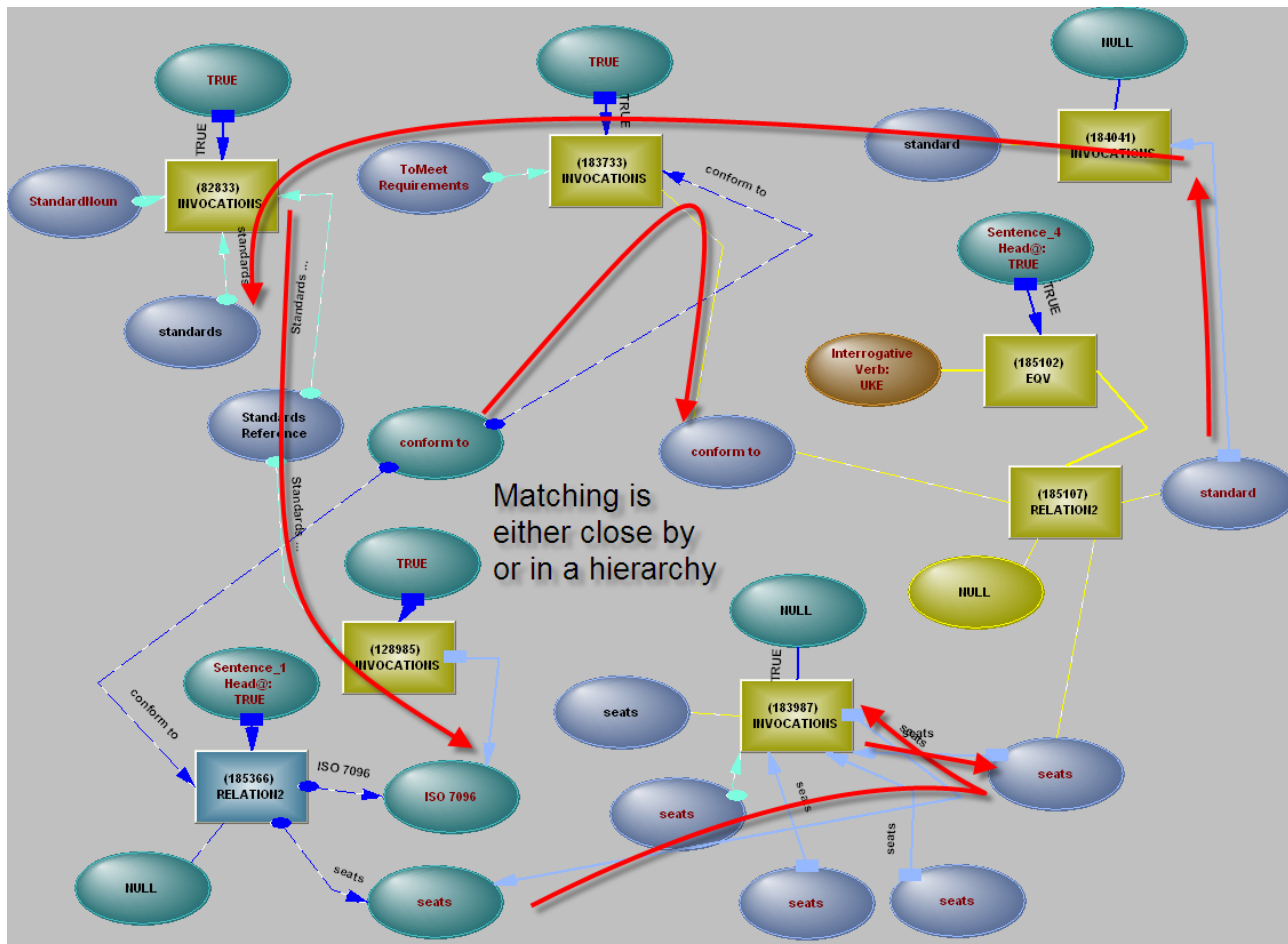


Building Semantic Structure

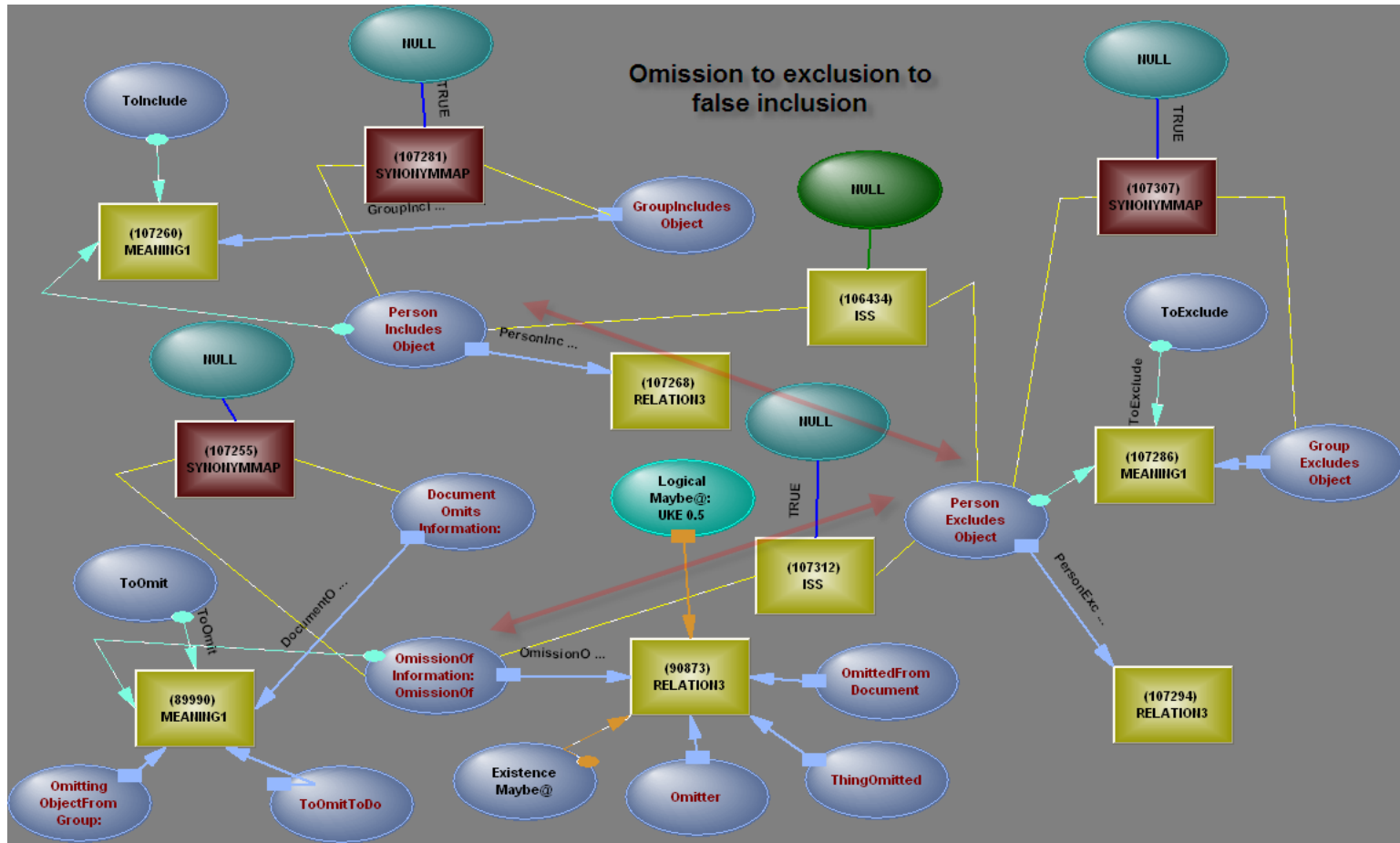


Searching

A search structure is created using free text, and it is then matched to the structure built from the document. These aren't two different structures – the query structure is anchored in the structure built from the text to be searched, so common points are found during building of the query, and defined terms are used.



Structural Synonyms



John omitted the data from the report
 The report omitted the data
 John didn't include the data in the report
 John left the data out of the report
 John forgot to put the data in the report

John failed to add the data to the report
 We were instructed not to insert the data
 John excluded the data from the report
 The report did not contain the data
 The contents of the report were incomplete

No Real Limit



There is no real limit on the extent of Semantic Search.

Currently the queries are separate entities, but they can easily be made to create objects or lists which are used in additional queries - the queries become part of an interactive dialogue with the machine, with both the machine and the user changing their internal states and connections.

Interactive Engineering Pty Ltd
Australia

www.inteng.com.au

ORION Technology