



# Linguamatics Update

Andrew Hinton

Chemaxon UGM

24<sup>th</sup> May 2016



# About Linguamatics



Software

Consulting

Hosted content



- Agile, scalable, real-time NLP-based text mining
- Fact extraction and knowledge synthesis

Pharma/Biotech

Healthcare

Government

Including 27 of  
the top 50

Including Kaiser  
Permanente

Including  
FDA, NIH

# Toolbox of Methods for Powerful Querying

## NLP

- Precise linguistic relationships, sentence co-occurrence
- Precise negation e.g. “pressure” but not “blood pressure”

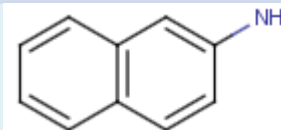
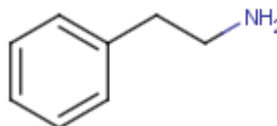
## Terminologies

- Search for concepts and classes, not just keywords
  - e.g. cancer and get synonyms **and** children:
  - Malignant neoplasms, Malignant tumor ...

## Regular Expressions

- Rule based pattern matching for e.g. measurements, lab codes, mutations
- e.g. microRNA: `let-?\d+.*` `mirn?a?-?\d+.*`

## Chemistry



## Fielded Search

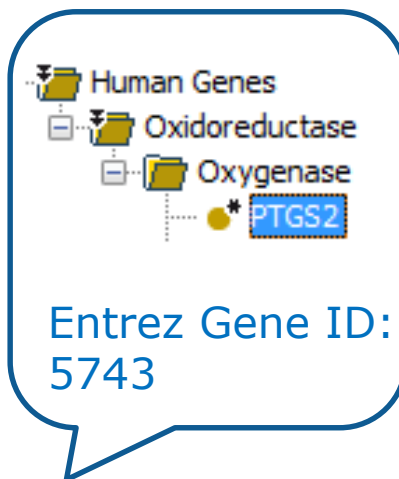
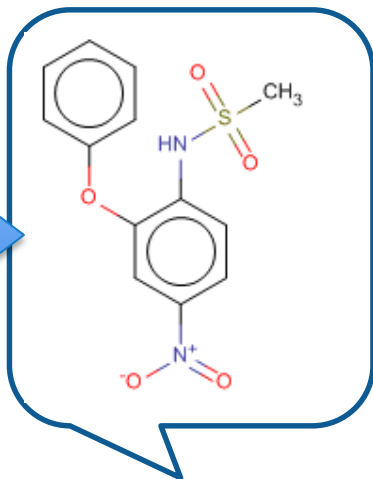
- Restrict within particular regions of a document, including nested e.g. table cell in table in Description

## High Throughput

- Simultaneous processing of large numbers of items e.g.
  - 500 compounds, 500 genes from microarray experiment, etc.

# From Words to Meaning

Chemical search embedded within NLP query



"Among them, nimesulide, a selective COX2 inhibitor, ..."

inhibits

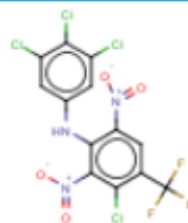
Linguistics to establish relationships



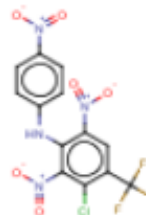
# Chemical Structures in Results Output

- ◆ Control over cell display in results output
- ◆ Chemical structures in html output

SupplierID	chemaxon
URL Prefix	/api?type=chemical_depiction/
URL Suffix	
Linkout Type	NodeID
HTML options	
Output type	<input type="radio"/> Linkout <input type="radio"/> Image <input checked="" type="radio"/> Custom template
<pre>&lt;img src="{{URL}}"/&gt; &lt;br&gt; {{PT_LINKOUT}}</pre>	
Template	



3-chloro-2,6-dinitro-N-(3,4,5-trichlorophenyl)-4-(trifluoromethyl)aniline

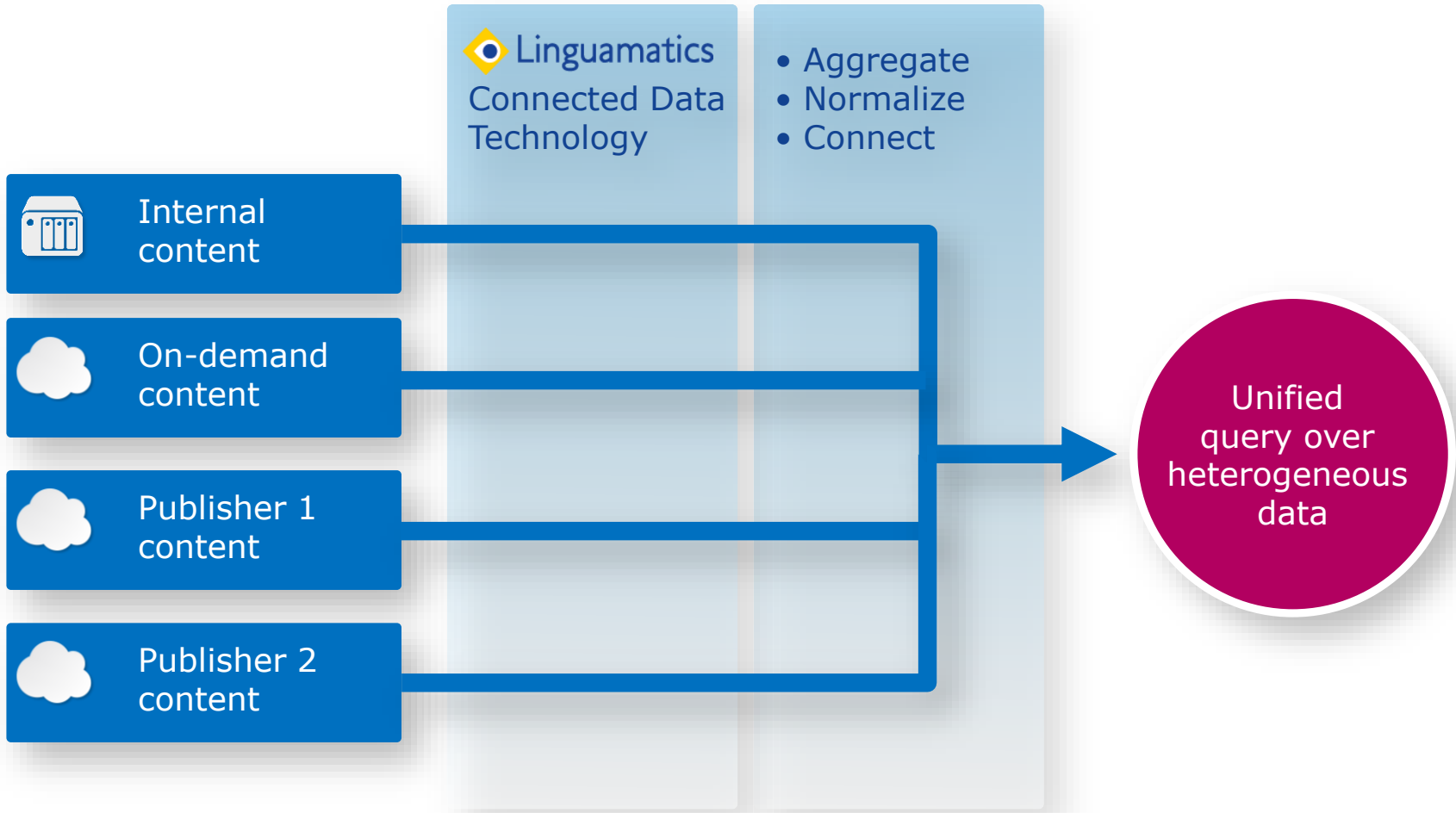


3-chloro-2,6-dinitro-N-(4-nitrophenyl)-4-(trifluoromethyl)aniline

# I2E Visualization



# Key I2E Focus: Data Integration



*"A couple of months in  
the laboratory can  
frequently save a couple  
of hours in the library"*  
Frank Westheimer

# Thank You

- [Healthcare Blog](#)
- [Life Sciences Blog](#)

For more information...

Visit: [www.linguamatics.com](http://www.linguamatics.com)

Contact: [Andrew.Hinton@linguamatics.com](mailto:Andrew.Hinton@linguamatics.com)