LinguaSys Confidential

LinguaSys

Understand, Analyze, Translate

Semantic NLU
The Missing LNKR™

Brian Garr

- CEO and founder LinguaSys, Inc.
- Managed IBM's Speech Recognition and MT Group
- Created WTS in 2000
- CTO of Globalink (an MT company) from 1995-1998
- 1998 Smithsonian Institution's "Heroes in Technology" award for his work in Machine Translation
- Chapter contributor; *Where Humans Meet Machines: Innovative Solutions for Knotty Natural-Language Problems*, edited by Amy Neustein and Judith A. Markowitz (Jun 30, 2013)

Characteristics of Words

- They have meaning(s)
- They may have gender
- They may be countable
- They may change form when pluralized
- They generally have a morphology
- They may change meaning when paired with other words
- They may describe animate or inanimate things
- If a Verb, is it transitive?
- ……

LinguaSys
Understand, Analyze, Translate

More about words

- Words have;
  - Synonyms = Car, Automobile
  - Homonyms = Tank (armed vehicle), Tank (blow a test), Tank (a container of gasoline)
  - Hyponyms = Tank<Armored Vehicles, Military Vehicles
  - Hypernyms = Tank>Panzer, Sherman
  - Domains = Toppings for Pizzas>Anchovies, Pepperoni, Meatballs….

Words in multiple languages

- Physical boundries? Not always
- Connection to the same concept in other languages?
  - One to One
  - One to Many
  - Many to Many
  - Many to One
  - Many to None
  - One to None

LinguaSys
Understand, Analyze, Translate

My definition of Semantics*

for purposes of this discussion

- The understanding of the concepts and ideas being presented in an uterance (textual or vocal) of words
  - What does the source intend it to mean?
    - Evaluate all of the words in the utterance
    - Look at the relationship to the other words in the utterance
    - Look for word phases
    - Disambiguate highly ambiguous words?

I need to fill the tank of the tank.

Words

- I
  - Segmentation#1
    - Morpheme#1: i
      - Guess#1: 1=NOUN,2=1,3=NO,4=PPS,5=ALL,9=1,15=NOM,2
        - Sense#1: [FamilyID#301, MapID#1]
- need
  - Segmentation#1
    - Morpheme#1: need
      - Guess#1: 1=VERB,2=ALL,3=ALL,5=ALL,8=REG,9=ALL,10=E
        - Sense#1: [FamilyID#112358, MapID#136357]
      - Guess#2: 1=VERB,2=ALL,3=ALL,5=ALL,8=REG,9=1,10=PRI
        - Sense#1: [FamilyID#112358, MapID#136357]
- to
  - Segmentation#1
    - Morpheme#1: to
      - Guess#1: 1=PREP,4=NA,15=PREP,20=TO,24=ALL,25=ACT
        - Sense#1: [FamilyID#323, MapID#452]
- fill
  - Segmentation#1
    - Morpheme#1: fill
      - Guess#1: 1=NOUN,2=1,3=YES,4=REG,5=ALL,9
        - Sense#1: [FamilyID#82204, MapID#87492]
      - Guess#2: 1=NOUN,2=1,3=YES,4=REG,5=ALL,9
        - Sense#1: [FamilyID#89237, MapID#95349]
      - Guess#3: 1=VERB,2=ALL,3=ALL,5=ALL,8=REG
        - Sense#1: [FamilyID#107857, MapID#126389]
      - Guess#4: 1=VERB,2=MANY,3=ALL,5=ALL,8=RE
        - Sense#1: [FamilyID#107857, MapID#126389]

.....Guess#38

Dependency Tree

I need to fill the tank of the tank

- I need [Sequence#1]
  - need to fill [Sequence#3]
    - to fill [Sequence#4]
      - fill tank [Sequence#5]
        - the tank [Sequence#6]
          - tank of the tank [Sequence#7]

LinguaSys
Understand, Analyze, Translate

The Missing LNKR

- A=1$T=301%A=136357$T=112358$Y13=MED%A=452$T=323%A=128399$T=108811$Y13=HIGH%A=117$T=309%A=3509686$T=28805$Y1=JAR$Y13=LOW%A=2761459$T=324%A=37343$T=34500$Y13=MED%

Language	Family ID	Lemma / Description	Note	Mapping
English	26925	tank		3006
English	34499	tank		3734
English	34500	tank	a military tank	3734
English	34503	tank		3734
English	83371	tank		8880
English	107475	tank	"tank animal refuse"	12555
English	117887	tank		14856
English	28805	tank	gas tank	350968

LinguaSys
Understand, Analyze, Translate

The Carabao Data Manager

Carabao Data Manager v2.9.1.0

Actions Data Reports Edit Window Help

Dictionary

Family ID | Text | Map ID | Last Update | Pragmatics.. | Links... | Thesaurus | Locate tank

Find substring | Move family | Bulk update | Bulk copy

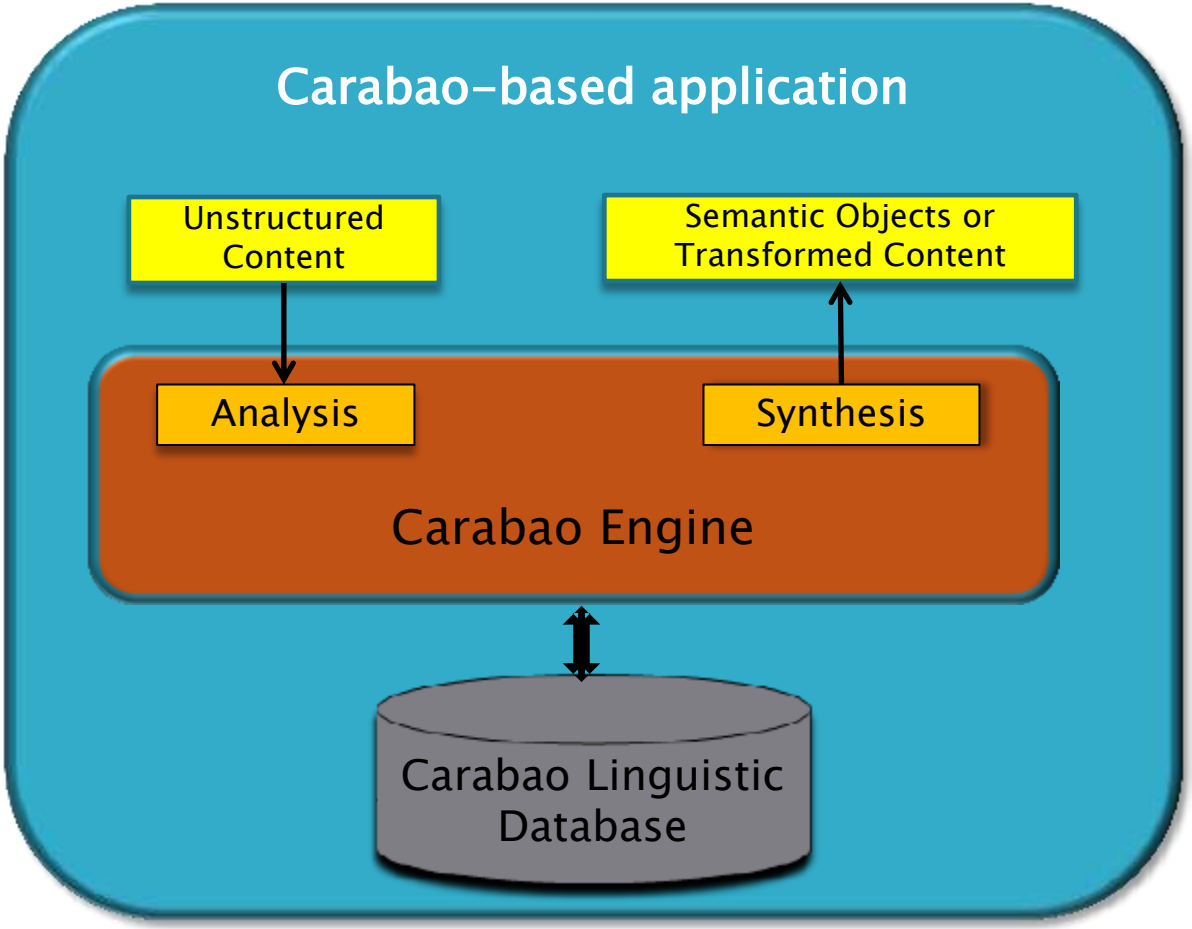
Language	Family ID	Lemma / Description	Note
English	26925	tank	
English	34499	tank	
English	34500	tank	a military tank
English	34503	tank	
English	83371	tank	
English	107475	tank	"tank animal refuse"
English	117887	tank	
English	28805	tank	gas tank
English	34503	tank car	tank car
English	34504	tank circuit	tank circuit
English	34505	tank destroyer	tank destroyer
English	67725	tank driver	tank driver
English	34506	tank engine	tank engine
English	56759	tank farm	tank farm

Close | Query | Copy | Insert | Chang

AR
CCC
CHI
DE
EL
✔ ENG
ES
FA
FI
FR
HE
ID
IPA
IT
JP
KO
MS
PS
PT
RUS
TH
TW
UR

Carabao Runtime Architecture

English
Spanish
Brazilian Portuguese
German,
Chinese (Simplified & Traditional)
Japanese
Thai
Vietnamese
Malay
Pashto
Russian
Persian
Urdu
Hebrew
Arabic
French



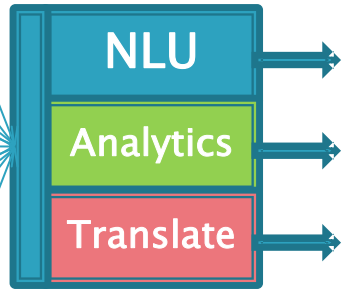
LinguaSys
Understand, Analyze, Translate

Carabao Engine

Language Neutral Engine		
Unique to LinguaSys. In the NLP space, not aware of any commercially available product that draws 100% of language processing information from the language models	Benefit: Cheaper and faster to add new languages and modifications No programming or application QA is required Work can be done by trained computational linguists, not software engineers	For Example: There is no incremental work to do to add new language support to an existing application, once the language model is finished

Language Neutral Knowledge Representation (LNKR)

Language	Input Sentence	Language-neutral representation
English	My cellular phone is perfect	A=22$T=301%A=177232$T=**26300**$Y10=2$Y12=FOSE$Y13=LOW%A=113$T=308%A=6117$T=**5309**$Y13=MED%
English	My mobile is perfect	A=22$T=301%A=2197726$T=**26300**$Y1=JAR$Y13=MED%A=113$T=308%A=6117$T=**5309**$Y13=MED%
Spanish	Mi móvil es perfecto	A=2082735$T=301%A=2197727$T=**26300**$Y1=JAR$Y13=MED%A=2082213$T=308%A=2091410$T=**5309**$R5=M$R25=ACT$Y13=LOW%
Russian	Мой мобильник идеален	A=2476403$T=114268%A=2903406$T=**26300**$Y1=TALK%A=14152983$T=**5309**%
S. Chinese	我的手机完美了	A=3078604$T=301$Y9=L%A=3225370$T=**26300**$Y13=LOW%A=3217686$T=**5309**$Y9=L%A=3199311$T=395$Y8=POS$Y10=1$Y13=LOW%



All applications support all languages immediately

LinguaSys
Understand, Analyze, Translate

LinguaSys Crosslingual Query Builder

English Search Query: Mexico Ice Murder

Spanish - español

Interpret

- Mexico
  - a republic in southern North America; became independent from Spain in 1810
    - México
    - Méjico
    - Mexico
- Ice
  - the frozen part of a body of water
    - hielo
    - helado
  - water frozen in the solid state.
    - hielo
  - a US agency in charge of immigration and customs
    - ICE
    - La Oficina de Aduanas y Control Migratorio
    - La Oficina de Aduanas y Control Migratorio de Estados Unidos
    - Servicio de Immigration y Aduanas
    - el Servicio de Immigration y Aduanas
  - put ice on or put on ice.
    - helar
    - hielo
    - hiele

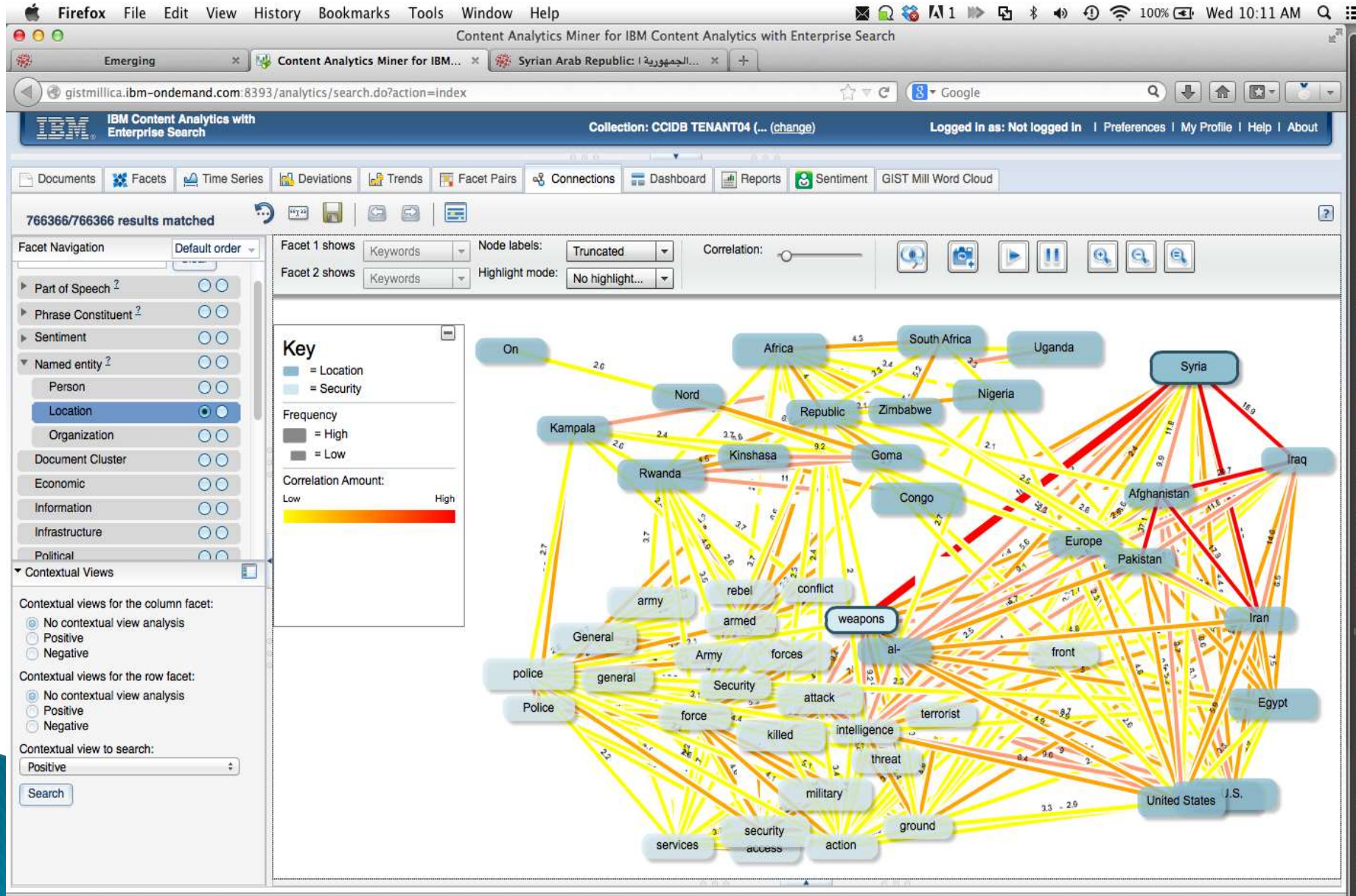
Multiple Spellings

Export to CSV

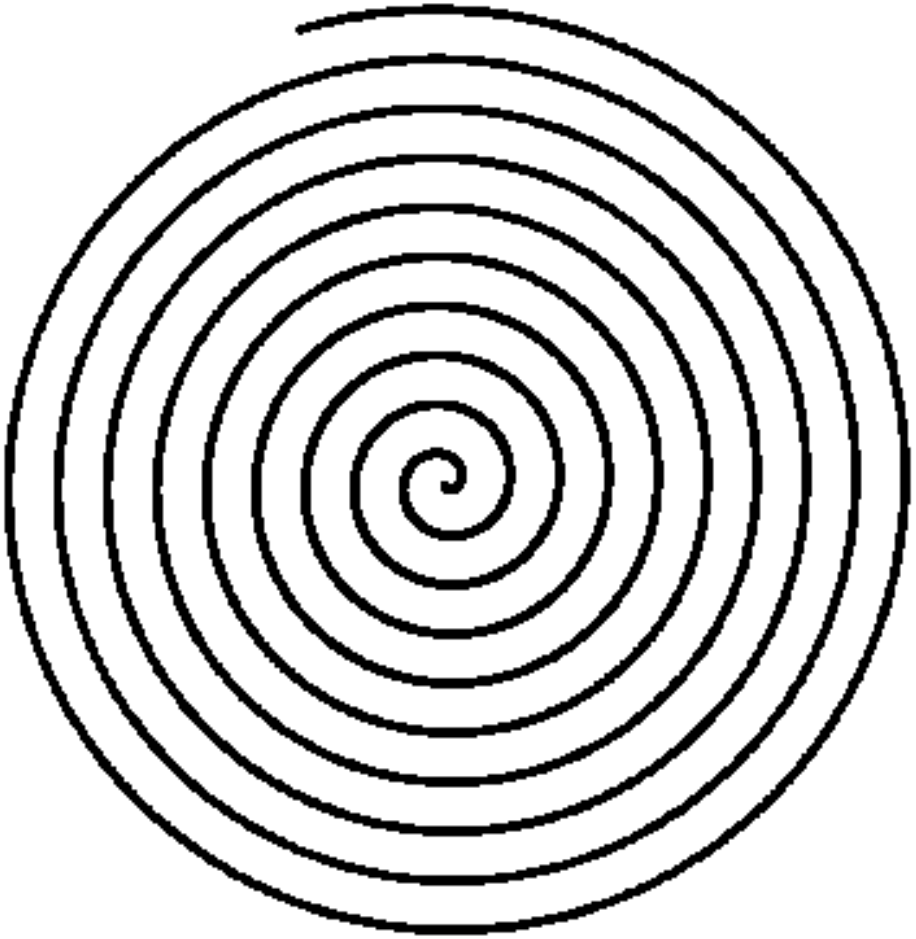
(México OR Méjico OR Mexico) (ICE OR La Oficina de Aduanas y Control Migratorio OR La Oficina de Aduanas y Control Migratorio de Estados Unidos OR Servicio de Immigration y Aduanas OR el Servicio de Immigration y Aduanas) (asesinato)

Language	Family ID	Lemma / Description
Español	57097	México
Español	57097	Méjico
Español	57097	Mexico

GIST MILL


Firefox File Edit View History Bookmarks Tools Window Help
Wed 10:11 AM
Content Analytics Miner for IBM Content Analytics with Enterprise Search
Emerging
Content Analytics Miner for IBM...
Syrian Arab Republic: الجمهورية...
gistmillica.ibm-ondemand.com:8393/analytics/search.do?action=index
IBM Content Analytics with Enterprise Search
Collection: CCIDB TENANT04 (... (change)
Logged in as: Not logged in | Preferences | My Profile | Help | About
Documents
Facets
Time Series
Deviations
Trends
Facet Pairs
Connections
Dashboard
Reports
Sentiment
GIST Mill Word Cloud
766366/766366 results matched
Facet Navigation
Default order
Part of Speech
Phrase Constituent
Sentiment
Named entity
Person
Location
Organization
Document Cluster
Economic
Information
Infrastructure
Political
Contextual Views
Contextual views for the column facet:
No contextual view analysis
Positive
Negative
Contextual views for the row facet:
No contextual view analysis
Positive
Negative
Contextual view to search:
Positive
Search
Facet 1 shows
Keywords
Facet 2 shows
Keywords
Node labels:
Truncated
Highlight mode:
No highlight...
Correlation:
Key
= Location
= Security
Frequency
= High
= Low
Correlation Amount:
Low
High
On
Africa
South Africa
Uganda
Syria
Nord
Republic
Zimbabwe
Nigeria
Kampala
Kinshasa
Goma
Rwanda
Congo
Iraq
Afghanistan
Europe
Pakistan
Iran
Egypt
United States
U.S.
rebel
conflict
army
armed
weapons
General
Army
forces
al-
front
police
general
Security
attack
Police
force
terrorist
killed
intelligence
threat
military
ground
security
services
action

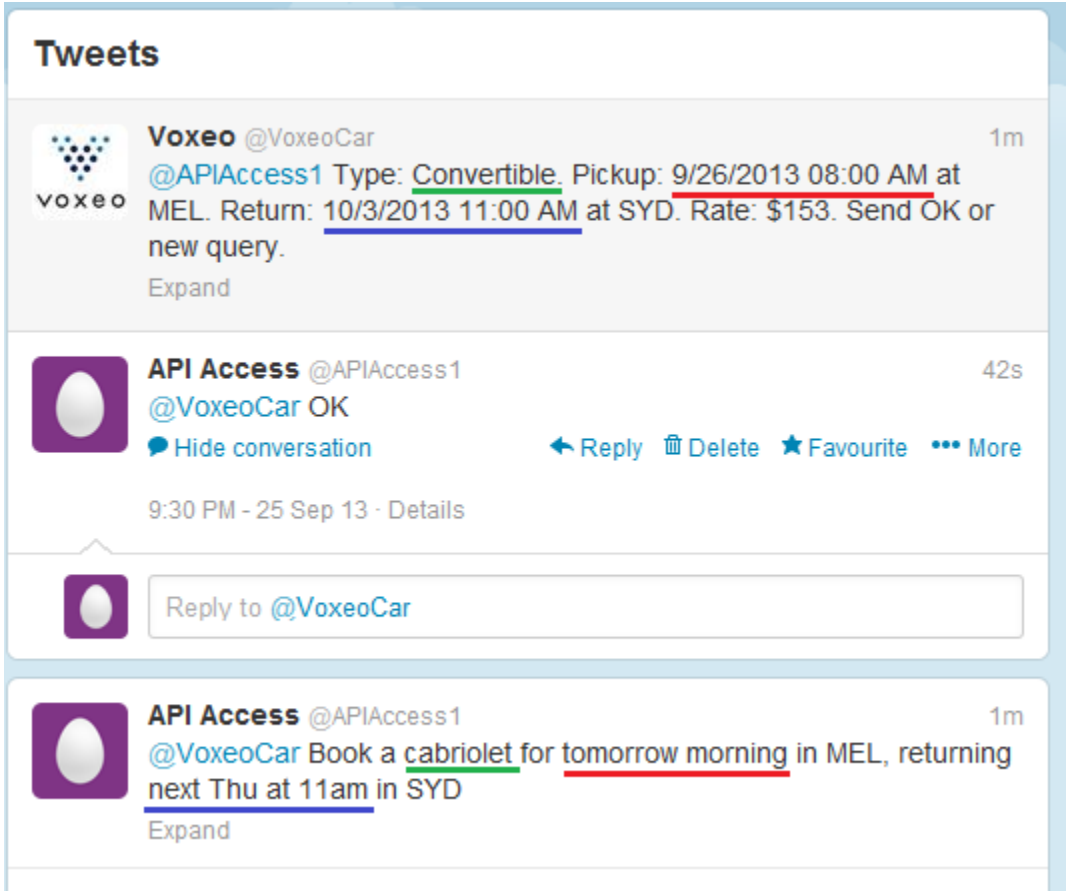
FORGET YOU SAW THAT CHART



LinguaSys
Understand, Analyze, Translate

Natural Language Understanding

- Can be hosted or embedded
- Understands 15 languages
  - 4 months to add new language
- No statistical models or embedded grammars
- Concept based
- Create new NLU apps in days, not months



LinguaSys
Understand, Analyze, Translate

NLU 2.0

- TEXT or SPEECH
- HOSTED or EMBEDDED
- 17 LANGUAGES
  - More each Q
- CONCEPT BASED NLU
- 10X DECREASE IN COST/SPEED TO BUILD

LinguaSys
Understand, Analyze, Translate

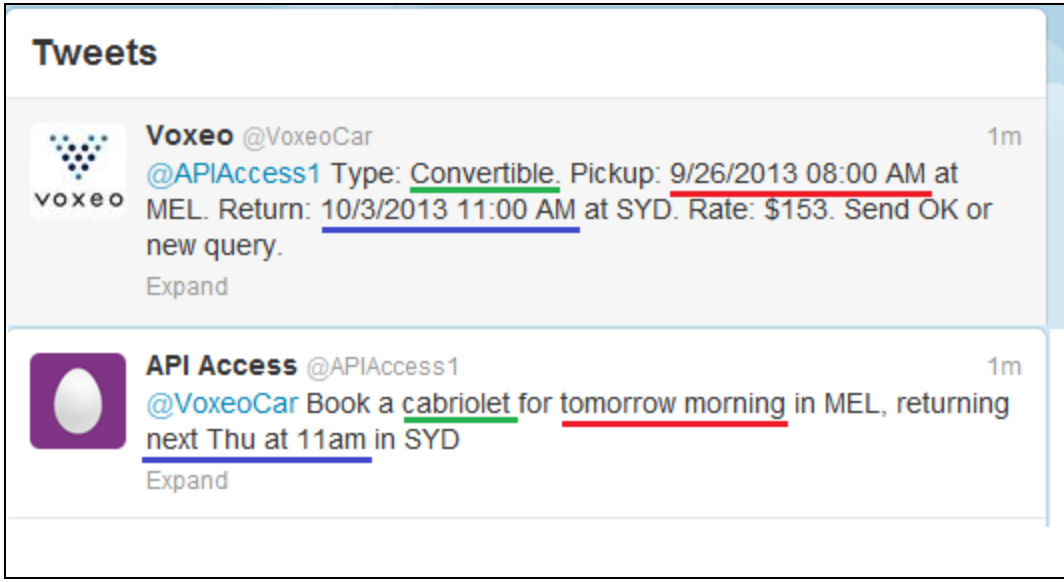
SEMANTIC UNDERSTANDING

- RESPONSE

Convertible,
9/26/2013,
10/3/2013

- REQUEST

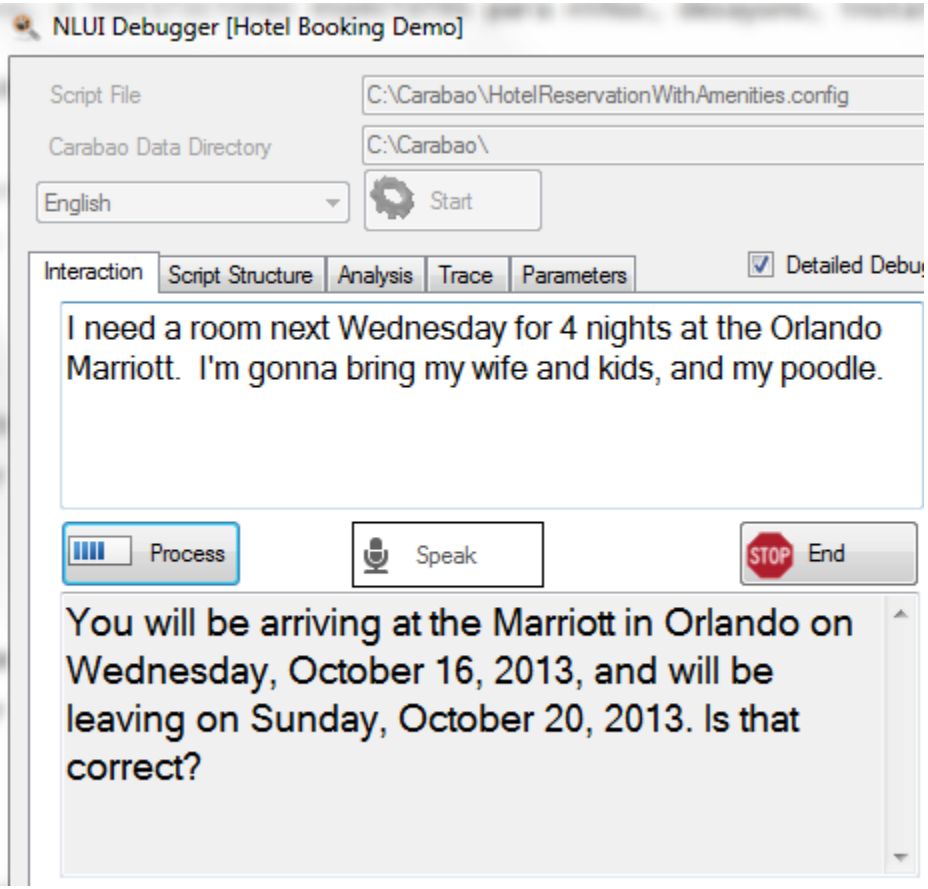
Cabriolet,
tomorrow,
next Thu



LinguaSys
Understand, Analyze, Translate

Hotel Booking

Today's date Tuesday, 10/08/13



- Accepts input in all supported languages
- Dialog management
- No SLMs
- No Embedded Grammars
  - But Out of Grammar messages
- Simple XML scripts
- Hotel Booking demo is 576 lines!

Analyze/Disambiguate every word/phrase

- I [detected as FamilyID#301, RecordID#1] at 0
- I need [FamilyID#1771, RecordID#3548675]
- need [detected as FamilyID#112358, RecordID#136357] at 2
- a [detected as FamilyID#310, RecordID#118] at 7
- room [detected as FamilyID#32808, RecordID#35864] at 9
- next [detected as FamilyID#95926, RecordID#103465] at 14
  - Clause: ["I need a room next Wednesday for 4 nights at the Orl
  - Sentence: ["I need a room next Wednesday for 4 nights at the
  - All interpretations
    - next [FamilyID#91982, RecordID#98070]
      - Grammar
      - Style
      - Hypernyms
    - next [FamilyID#95926, RecordID#103465]
    - next [FamilyID#97332, RecordID#106202]
    - next [FamilyID#100859, RecordID#112883]
  - Grammar
  - Hypernyms
    - [FamilyID#95926] next
    - [FamilyID#102706] coming
    - [FamilyID#5280] future
- Wednesday [detected as FamilyID#90929, RecordID#96861] at 19
- for [detected as FamilyID#440, RecordID#1073] at 29
- 4 [detected as FamilyID#1999999999, RecordID#-28] at 33
- 4 nights [FamilyID#1101, RecordID#12523308]
- 4 nights [FamilyID#1101, RecordID#2179234]
- 4 nights [FamilyID#1101, RecordID#402012]
- 4 nights [FamilyID#1101, RecordID#1219]
- nights [detected as FamilyID#90886, RecordID#96832] at 35
- at [detected as FamilyID#429, RecordID#241392] at 42
- the [detected as FamilyID#309, RecordID#117] at 45
- Orlando [detected as FamilyID#58564, RecordID#64156] at 49
- Marriott [detected as FamilyID#750474, RecordID#278644] at 57

Language	Family ID	Lemma / Description	Note
English	91982	next	"next the doctor examined his back"
English	95926	next	"the following day"; "next in line"; "th
English	97332	next	"had adjacent rooms"; "in the next r
English	100859	next	"our next president"

- próximo [detected as FamilyID#95926, RecordID#2088372] at 27
- miércoles [detected as FamilyID#90929, RecordID#2103089] at 35
- por [detected as FamilyID#440, RecordID#2082631] at 45
- 4 [detected as FamilyID#1999999994, RecordID#-285] at 49
- 4 noches [FamilyID#1101, RecordID#2179274]
- 4 noches [FamilyID#1101, RecordID#2176516]
- noches [detected as FamilyID#90946, RecordID#2103102] at 51
- en [detected as FamilyID#431, RecordID#2082469] at 58
- el [detected as FamilyID#309, RecordID#2023781] at 61
- Marriott [detected as FamilyID#138148, RecordID#2933774] at 64
- Orlando [detected as FamilyID#58564, RecordID#12938514] at 73

Dictionary

Changing an existing dictionary entry

Family ID: 1101 New Same As...

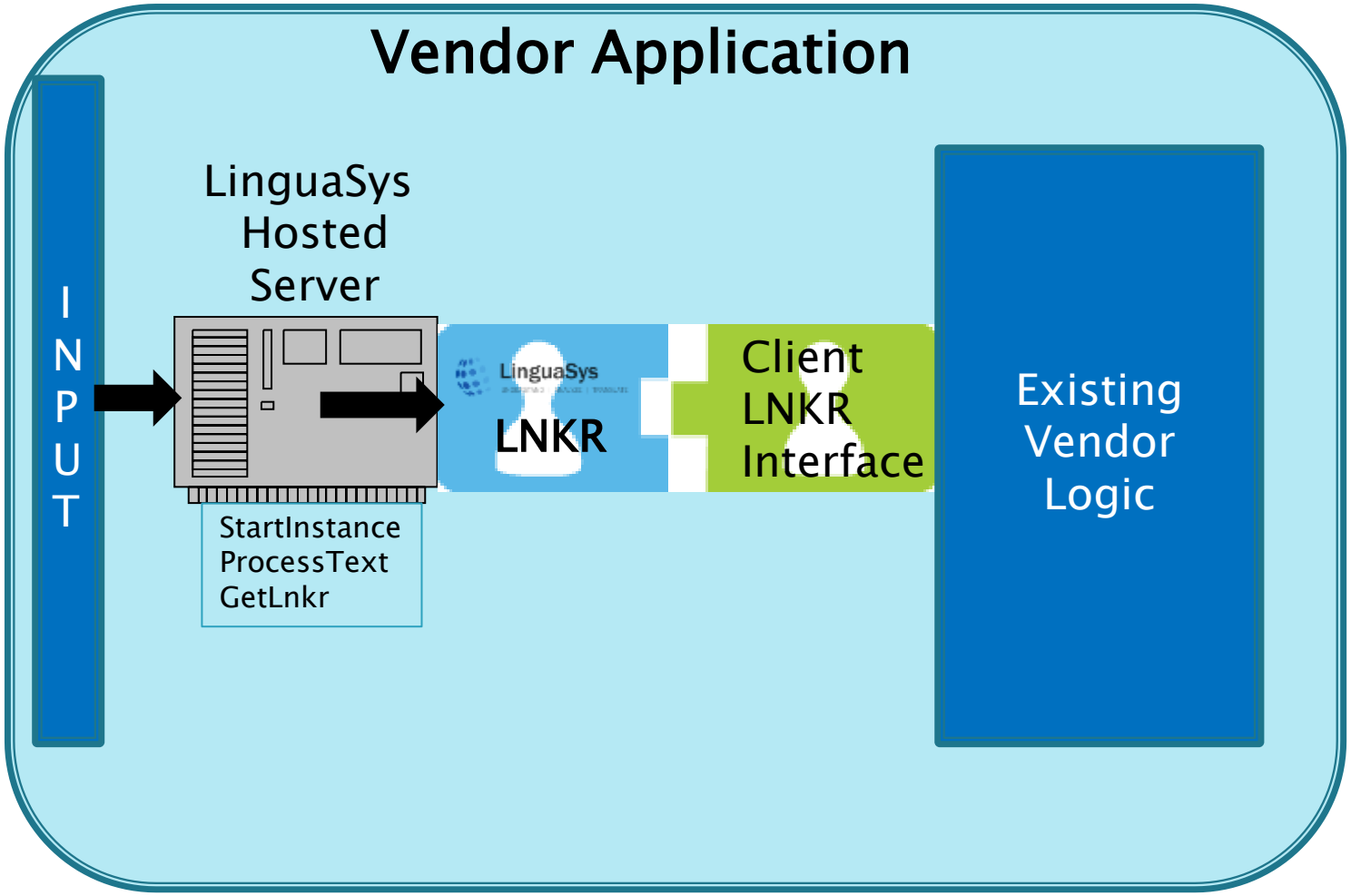
Language: ENG ... English

Edit Sequence

number, identity 1
category: large integer, identity 4
number, identity 3
(possible gap)
allowed: regular noun
allowed: adjective
(possible gap)
plural,countable,regular noun, priority 1 (head), identity 2
not allowed:regular noun

LinguaSys Translate

English,
Arabic,
Traditional Chinese,
Simplified Chinese,
German,
French,
Hebrew,
Japanese,
Malay,
Spanish,
Pashto,
Persian,
Portuguese,
Russian,
Thai,
Vietnamese,
Urdu



LinguaSys
Understand, Analyze, Translate

LinguaSys Value

- Faster, Cheaper, and Better
- Text and Speech Input
- Multiple languages

LinguaSys
Understanding@Work

info@linguasys.com
561-755-7150