Human Language Technology

Center for Content Extraction

Content Extraction Analytics
SIGDEV *End-to-End* Demo

[redacted]

21 May 2009

Derived From: NSA/CSSM 1-52
Dated: 20070108
Declassify On: 20330108

Introduction to Content Extraction

- New technologies can find Essential Elements of Information in documents
- The Center for Content Extraction provides “one stop shopping” for these technologies at NSA

Extraction can benefit SIGDEV from end to end

- Selection
- Translation & Transliteration
- Analysis
- Interpretation/Enrichment
- Retrieval
- Storage & Distribution

STAIRS Partners

S (Marina, CEA)

T (Cybertrans)

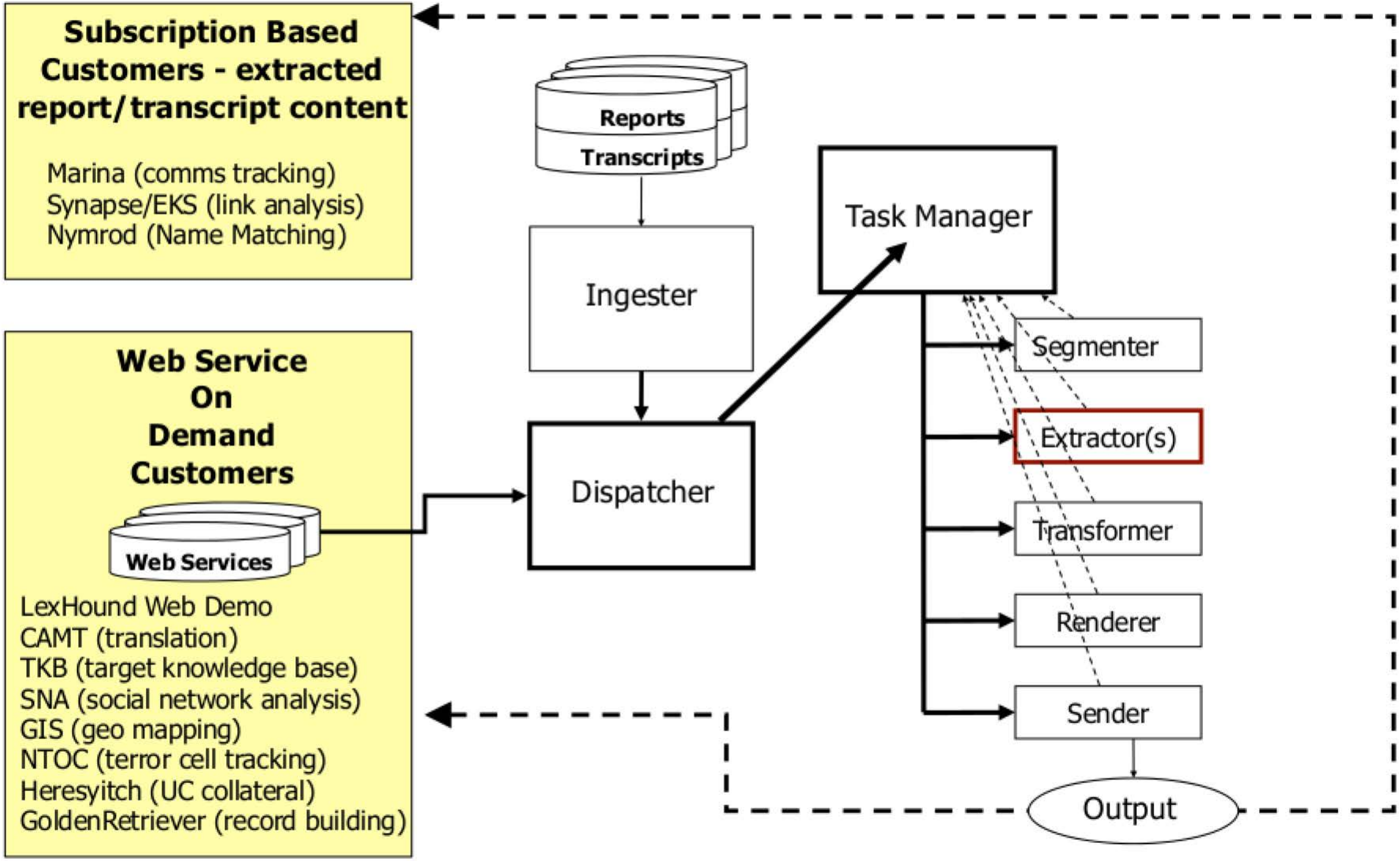
A (SNA/Paintball, Synapse)

I (Nymrod,Thundercloud)

R (Journeyman/CPE)

S (GoldenRetriever, SocioPath)

Implementation: CCE Extraction Architecture (LexHound)



Subscription Based Customers - extracted report/transcript content

Marina (comms tracking)
Synapse/EKS (link analysis)
Nymrod (Name Matching)

Web Service On Demand Customers

LexHound Web Demo
CAMT (translation)
TKB (target knowledge base)
SNA (social network analysis)
GIS (geo mapping)
NTOC (terror cell tracking)
Heresyitch (UC collateral)
GoldenRetriever (record building)

Elaboration: *The Central Importance of Storage*

- Each of the STAIRS Steps exploits stored information
  - Selection Dictionaries (“get it”)
  - Linguistic Glossaries for Translation
  - Wikis etc for enrichment (“know it”)
- Manual record-formation is slow, prone to omissions and inconsistencies
  - <200K Person Targets in TKB
  - Growth ~= 20K/year
- Automatic extraction accelerates storage
  - >3000K Citation Records in ***Nymrod*** Entity DB
  - Growth ~= 1000K/year

Machine vs. Manual Chief-of-State Citations

	Nymrod (machine-extracted) Citations				**Last TKB Manual Update**
	Name	**Role**	**Code**	**Cites**	
1	Abdullah Badawi	Malaysian Prime Minister	cos	> 100	10/15/2007
2	Abdullahi Yusuf	Somali President	cos	> 300	N/A
3	Abu Mazin	(Mahmud 'Abbas) PA President	cos	> 200	5/20/2009
4	Alan Garcia	Peruvian President	cos	> 100	N/A
5	Aleksandr Lukashenko	Belarusian President	cos	> 50	N/A
6	Alvaro Colom	Guatemalan President	cos	> 200	N/A
7	Alvaro Uribe	Colombian President	cos	> 700	N/A
8	Amadou Toumani Toure	Malian President	cos	> 50	N/A
9	Angela Merkel	German Chancellor	cos	> 300	N/A
10	Bashar al-Asad	Syrian President	cos	> 800	N/A
...	....	....	...		
122	Yuliya Tymoshenko	Ukrainian Prime	cos	> 200	N/A