



Basis Technology's Just Right Solutions: Bigger than a Component, Smaller than a Stovepipe

Benson Margulies
Chief Technology Officer
June 7, 2007

Basis Technology Corporation

P 617.386.2000
800.697.2062 (toll-free)
F 617.386.2020
W info@basistech.com
www.basistech.com



What's New at Basis

- Farsi and Urdu Base Linguistics
- Entity Extraction Improvements
- Rosette Name Translator
- Rosette Name Search Engine
- Digital Forensics Products
- Arabic Desktop Suite



The Language Problem

- We've all been *here* before
- Much critical information in foreign language
- Linguists are scarce and precious
- Disarray of foreign names in Latin alphabet



Components Available Today

- *Extracting names*
 - Addresses critical information in foreign languages
- *Cross-language matching of names*
 - Extends existing correlation and matching to foreign language
 - Addresses disarray of foreign names in Latin alphabet
- *Translating names into English*
 - Semi- to fully-automatic transliteration and translation
 - Addresses scarcity of linguists



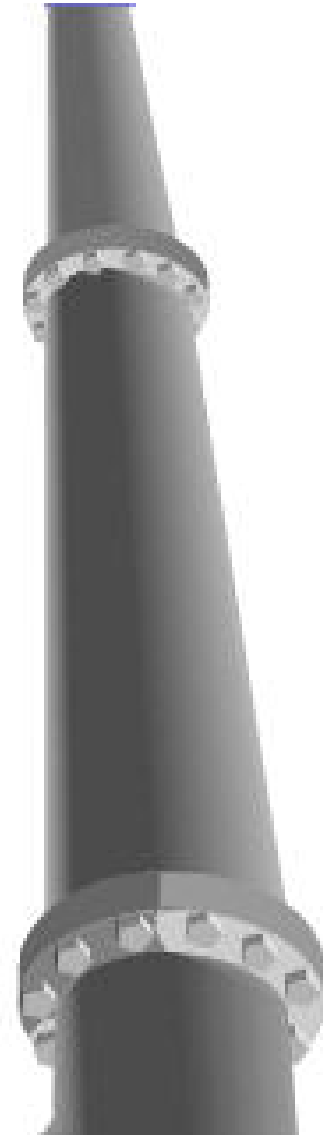
“Components Are Not Enough”

- Customers and partners spoke.
- We acted.
- Solutions in 2006 for enterprises and individuals.



End-User's Dilemma: Books and Stovepipes

- Analysts and linguists have tools that don't work together
- Needed: critical linguistic tools to cut across all these systems
- Accessible in HQ or in the field





Helping End-Users: Arabic Desktop Suite

- Simple plug-ins for Microsoft Office and Windows
 - Rapid name translation
 - Dictionary and gazetteer lookup
 - Digital maps and geo-location
- Next steps
 - More languages (Persian, Urdu...)
 - More features
 - More integration

Field intel report
...located at Safeer al feel ...

1. Fuzzy name match

Select Features

Fuzzy Search: safeer al feel

Search Clear

Detail Names

Go Help

Ortho match Phono match Limit to view

Nahr Safir al-Fil نهر صفير الفيل CNLQ 35°11' N 43°52' E

2. Link with geospatial data

US Board on Geographic Names (BGN)

Nahr Şafir Al Fil

Latitude: 35°10'34" N Longitude: 43°51'41" E

UFI: -3110217 JOG: NI38-02 FC: H DSG: CNLQ abandoned canal

Native script: Vocalized

نهر صفير الفيل

Feature Details

US Board on Geographic Names (BGN)

Nahr Şafîr Al Fîl

Latitude: 35°10'34" N Longitude: 43°51'41" E

UFI: -3110217 JOG: NI38-02

Native script:

نهر صفير الفيل

OK
Names
Copy
Help

3. Integrates names with geodata to return meaningful results, finds original Arabic.

Iraq - GeoScope by Basis Technology

File Edit Calibration View Help

Zoom 70% Detail... Go->

Latitude: 35°11' N Longitude: 43°52' E

نهر صفير الفيل Nahr Şafîr Al Fîl

Latitude 35°48' N Longitude 46°10' E 385 PE 0578 6200 Gazetteer Calibrated NUM



Helping System Builders

- Common shared tasks:
 - Ingest documents, find names, translate names
 - Store names
 - Look up names across scripts and languages
 - Correlate documents via names
- From components to systems
 - Solving end-to-end problems: Rosette Name Search Engine, Extending RLP to names and documents

Rosette Name Search Engine (RNSE)

- From comparing names across scripts
- To building persistent indices
- Example: Search Arabic Wikipedia in Chinese Pinyin

Enter query:

Index:

Minimum Match Score (0-1.0):

Maximum Results to Return:

Select Language:

Name to search for:

Score	Relevance	Name	Info
44%	<div style="width: 44%;"></div>	ماو زيدونغ	Maw Zydwngh

ماو تسي تونغ

من ويكيبيديا، الموسوعة الحرة

ماو تسي تونغ (24 ديسمبر 1893 إلى 9 سبتمبر 1976) - زعيم الحزب الشيوعي الصيني منذ 1955 حتى وفاته. تزوج مهنياً أمام الجيش الوطني في "الدميرة الحمراء". وعضو تأسس الحزب الشيوعي حتى كتابة هذا المقال. حافظ الحزب على قيادة الصين نتيجة انتصار ماو تسي تونغ في الحرب الأهلية في الصين وتأسيس جمهورية الصين الشعبية عام 1949.

عدل ماو على تقرير ملهون بنابند الموسوعة ششي "بالخطأ" وكان حرج

ويكيبيديا الموسوعة الحرة

- الصفحة الرئيسية
- الأحداث الجارية
- أحداث التاريخ
- أحداث الفترات الأساسية
- صاحب
- التاريخ
- أحدث

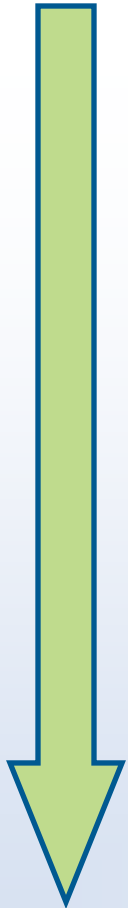


Extending the Pipeline Beyond Text

- Processing plain text is good ...
- Processing documents and names is better ...
- Documents:
 - Deconstructing structure
 - Running extraction
 - Detecting duplicated names
 - Annotating original document
- Names: Translation \Rightarrow Lookup \Rightarrow Storage



For Example ...



- Analyze HTML
- Extract entities
- Detect duplicate names
- Look up names in gazetteers or watch lists
- Translate names
- Store names in an index
- Annotated user friendly output of original text



Next Steps

- Combine end-user focus with enterprise resources
- Bring enterprise power to the desktop
 - If you can look-up names in a gazetteer on your desktop, why not all the gazetteers on your network?
- Stay tuned for future announcements!