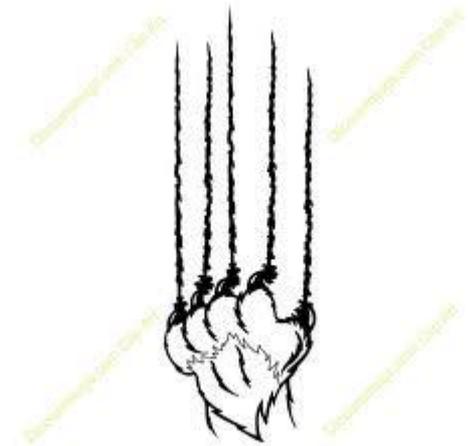


Sharpening the Claws on CAT tools



Increase Quality & Production,
Maximize Limited Resources



- The Linguist
- Computer Assisted Translation Tools
- Return on Investment
- Finally, delve into a short demo

The Problem

- Cleared linguists are a **scarce** and **costly** commodity
- Many linguists and analysts rely on **rote memorization** of the IC standards which may result in errors

The Approach

- Working in concert with **Computer-Assisted Translation** tools improves efficiency and accuracy
- CAT tools significantly accelerate **production**

Without CAT



The US Intelligence Community (IC) Transliteration Standard F

| Standard | Vowels (1) | Double Consonants | Teh Marbuta ّ (U+0629) | Digraphs (2) |
|---------------------------|------------|--|------------------------|--------------|
| Arabic IC Standard | N/A | not digraphs | "t" or "h" | No |
| Arabic (FBIS) | N/A | All letters except: - "ay" (end of word) - "ay" (elsewhere) | "t" (construct state) | No |
| Persian-Farsi IC Standard | N/A | - No digraph - to reflect sun letter (like "Shamsoddin" not "Shams al-Din") | n/a | No |

IC does not distinguish between long and short vowels.
Double consonants represented by the Arabic shaddah are shown in most cases (Hassan, Muna).
Exceptions: "ay" and consonants represented by digraphs are not doubled (al-Qadiri, al-Qadiri).

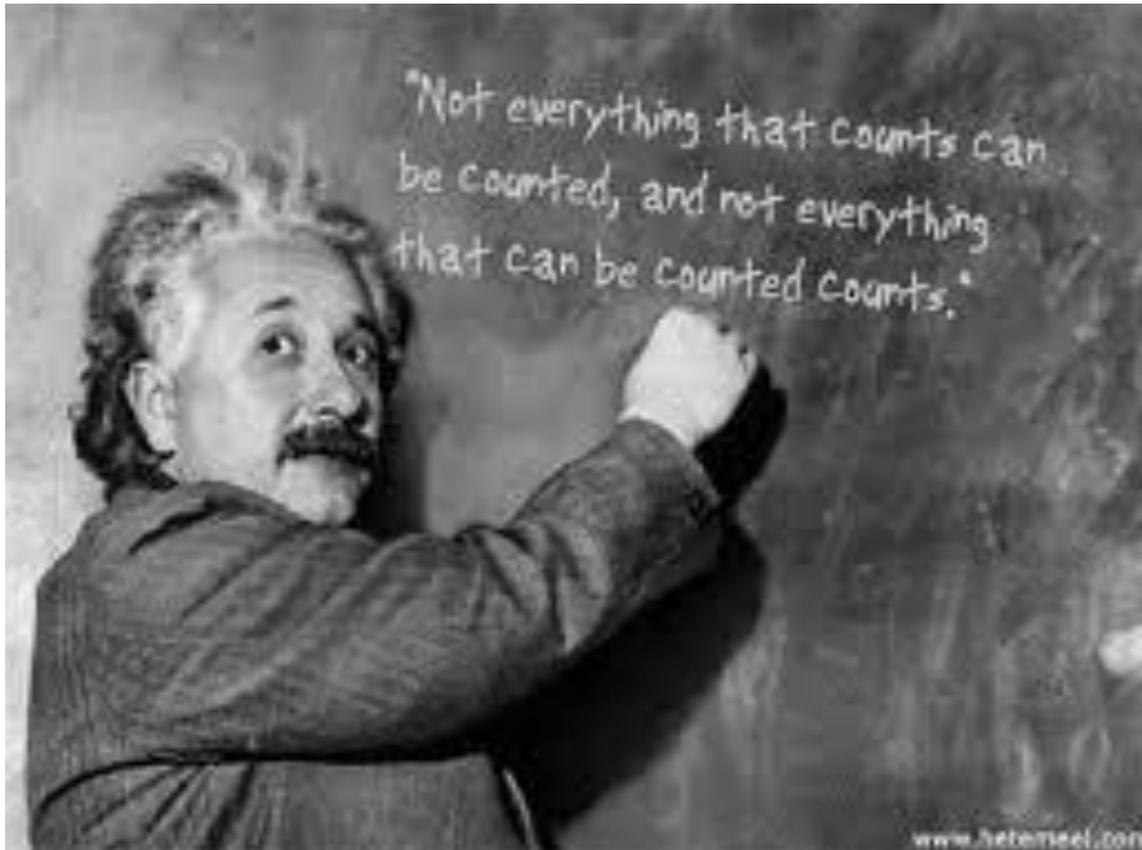
= 10 Seconds per word



©Ron Leishman * illustrationsOf.com/1082911

= 3 Seconds per word

What is ROI



ADDING CAT CAPABILITIES
IS APPROXIMATELY 28 % OF AN EMPLOYEE.
TOGETHER, EFFICIENCY MAY BE INCREASED BY AS MUCH AS
TWO-THIRDS!

LINGUIST

THE LINGUIST TRANSLATES
THE 10,000 NAME PHONE BOOK

TIME = 80+H WITH NO BREAKS
(MORE WITH QA)

COST = \$2,548 (\$91.00 PER
HOUR FOR THE LINGUIST)

ERROR RATE DEPENDS ON THE
LINGUIST

HIGHLIGHT + LINGUIST

HIGHLIGHT AND LINGUIST
TRANSLATE THE SAME PHONE BOOK

TIME = LESS THAN 8H

COST = \$733 (\$91.61 PER HOUR
BASED ON LICENSE COSTS AND THE
LINGUIST)

ERROR RATE = LESS THAN LINGUIST
ALONE

IN AUTOMATIC MODE, HIGHLIGHT SPEEDS ARE MUCH FASTER

- **Interactive** and automatic translation
- Standardization to the **IC Standard** and others
- **Dictionary** look up
- Access to the CIA's **Chiefs of State**
- Ability to **type** in Arabic or Persian from an English keyboard
- Quickly look up **foreign features** on a map

Farid Mustafa
Fered
Mustaffa
Moustapha
Mostafa
Mostapha
Fareed
F Moustaffa
F Moustafa
Fared

- **Transliteration Assistant**
 - Standardize names of people and places
 - Automatically translate lists of names
 - Dictionary and Gazetteer look up
- **Arabic Editor**
 - Compose and edit Arabic and Persian documents using a Western keyboard
- **Geoscope Map Viewer**
 - Rapid access to maps of the Middle East
 - Multilingual place name lookup

The Problem

- Cleared linguists are a **scarce** and **costly** commodity
- Many linguists and analysts rely on **rote memorization** of the IC standards which may result in errors

The Approach

- Working in concert with **Computer-Assisted Translation** tools improves efficiency and accuracy
- CAT tools significantly accelerate **production**

How CAT can help

- **Interactively** translates names of people and places between foreign languages and English
- **Automatically** translates large lists of names, such as telephone books, mobile device contacts, personnel rosters
- **Standardizes** names according to agency-approved systems
- Arabic/English dictionary of **common** terms
- **Place** name search
- Arabic/English dictionaries of **specialized** terms (weapons, IED, etc.)
- Country gazetteers for **place names** lookup
- CIA **World Leaders** list

The Story of Moustafa, the Soccer Spy

In **Kerbala**, **Moustafa Fared**, abandoned the car and the British uniform he'd worn for safety and walked to an American base near **Bucca**. Posing as a stranded motorist, **Farid** was given water, lunch, and a ride to the train station.

Captain **Samir Khaleel**, drove from Kabul to meet with **Fared** at the train station. He brought along a list of potential soccer players that might join the team.

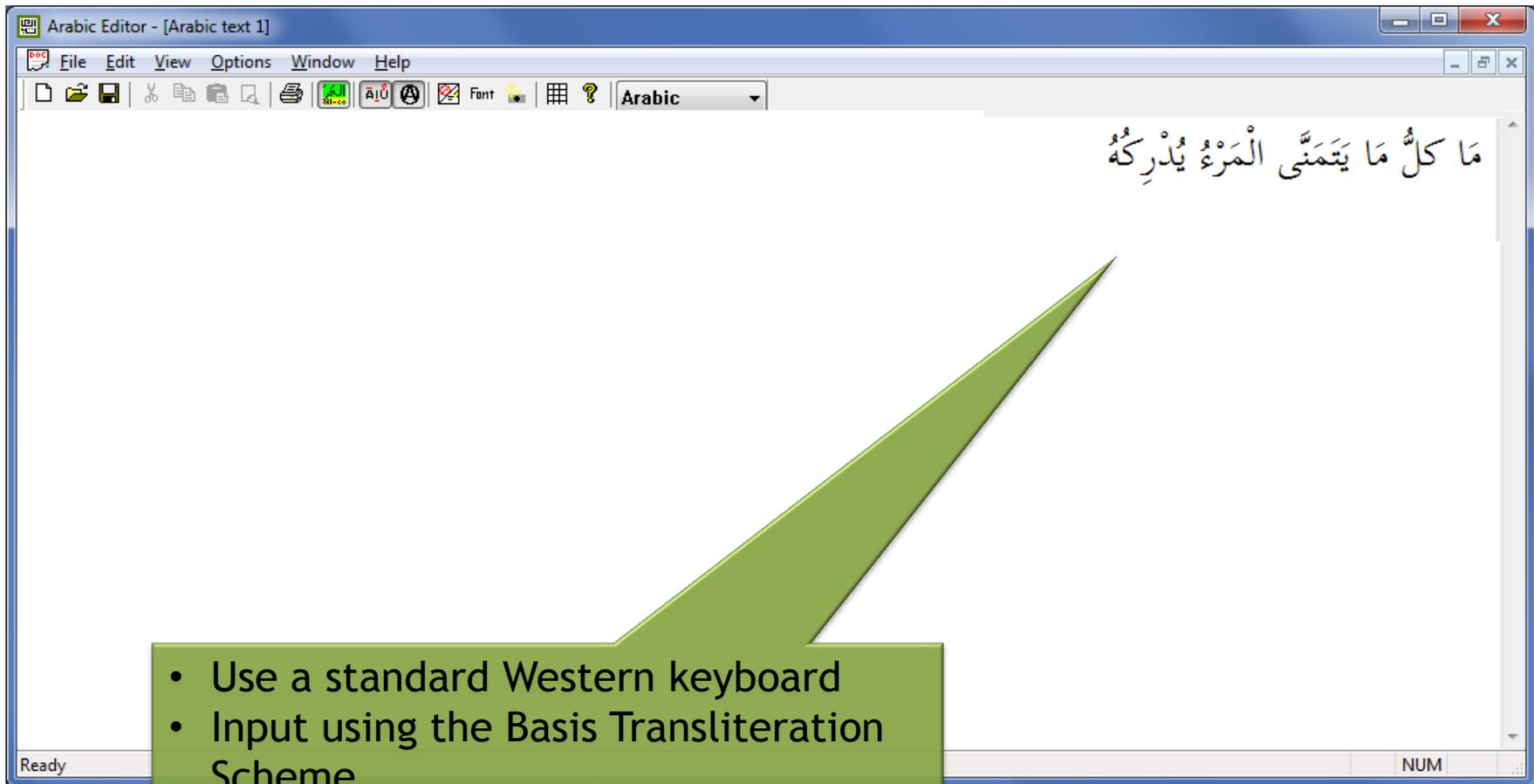
| | |
|---|----------|
| 1 | عمرخان |
| 2 | محمدنعى |
| 3 | نادرخان |
| 4 | محمدصديق |
| 5 | محمدزمان |
| 6 | شيررحمن |
| 7 | نادرشاه |
| 8 | محمدآغا |

تخلى **مصطفى فريد** وهو في **كربلاء** لسبب أمني عن سيارته والزي العسكري البريطاني الذي كان يرتديه وسار إلى قاعدة أمريكية قرب بوكا. متنكر بزي سائق سيارة عادي، أعطي فريد المياء والغداء وسيق إلى محطة القطار.

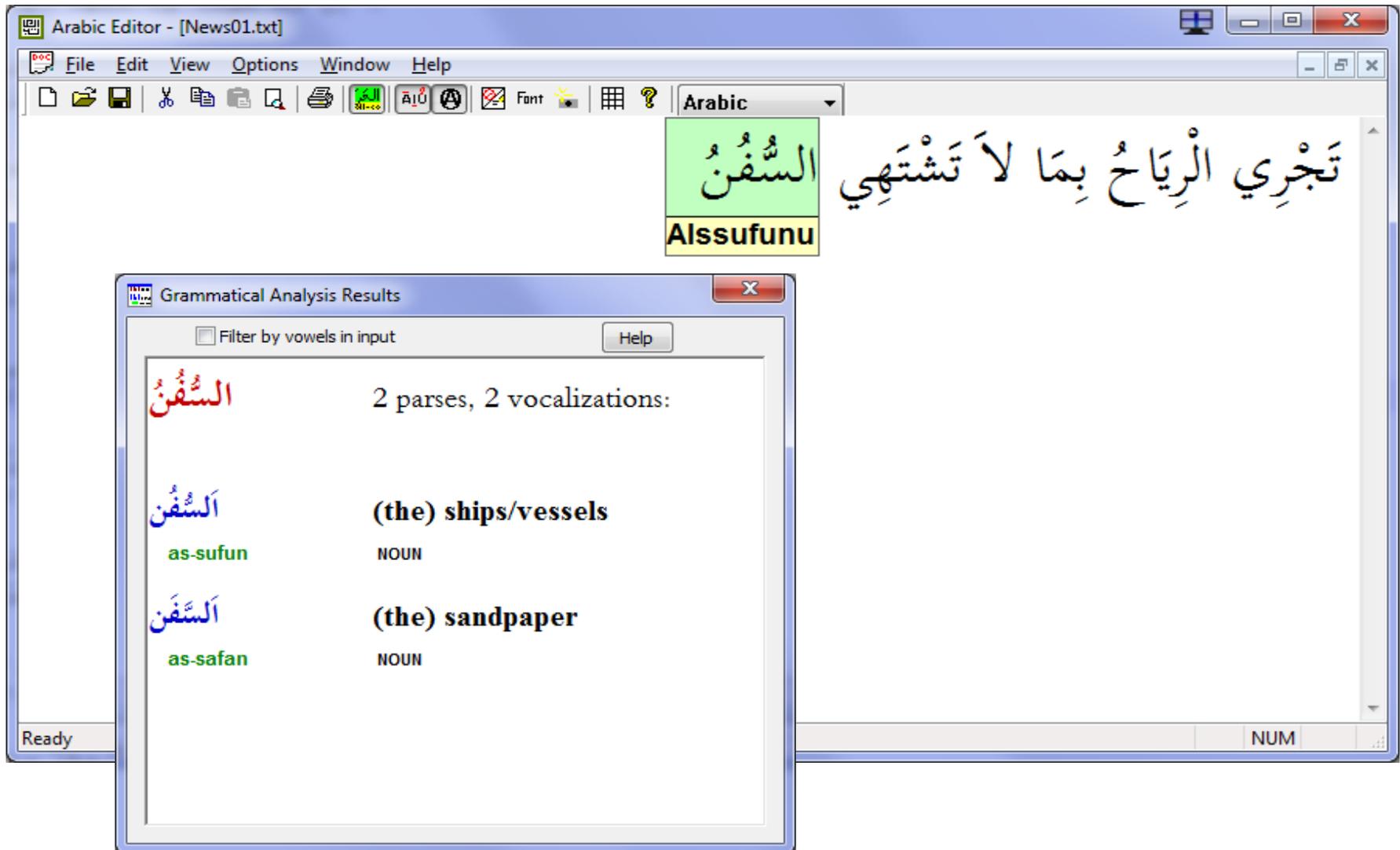
جاء الكابتن سمير خليل من كابول للاجتماع مع فريد في محطة القطار. وأحضر معه قائمة بأسماء لاعبي كرة القدم المحتمل إنضمامهم إلى الفريق.

| | |
|---|----------|
| 1 | عمرخان |
| 2 | محمدنعى |
| 3 | نادرخان |
| 4 | محمدصديق |
| 5 | محمدزمان |
| 6 | شيررحمن |
| 7 | نادرشاه |
| 8 | محمدآغا |

- Advanced input system for vocalized Arabic using standard keyboard
- Basic input system for Persian (Farsi/Dari) using standard keyboard
- Arabic/English dictionary for reading and analyzing Arabic text
- Phonetic “fuzzy” search capability



- Use a standard Western keyboard
- Input using the Basis Transliteration Scheme
- Easily learned within an hour
- Completely reversible Arabic-to-English system



The screenshot displays the Arabic Editor application window titled "Arabic Editor - [News01.txt]". The main text area contains the Arabic phrase "تَجْرِي الرِّيحُ بِمَا لَا تَشْتَهِي السُّفُنُ" (The wind blows by what the ships do not desire). A green box highlights the word "السُّفُنُ" (Alssufunu), which is also shown in a yellow box below it. A "Grammatical Analysis Results" window is open in the foreground, displaying the analysis for the word "السُّفُنُ".

Grammatical Analysis Results

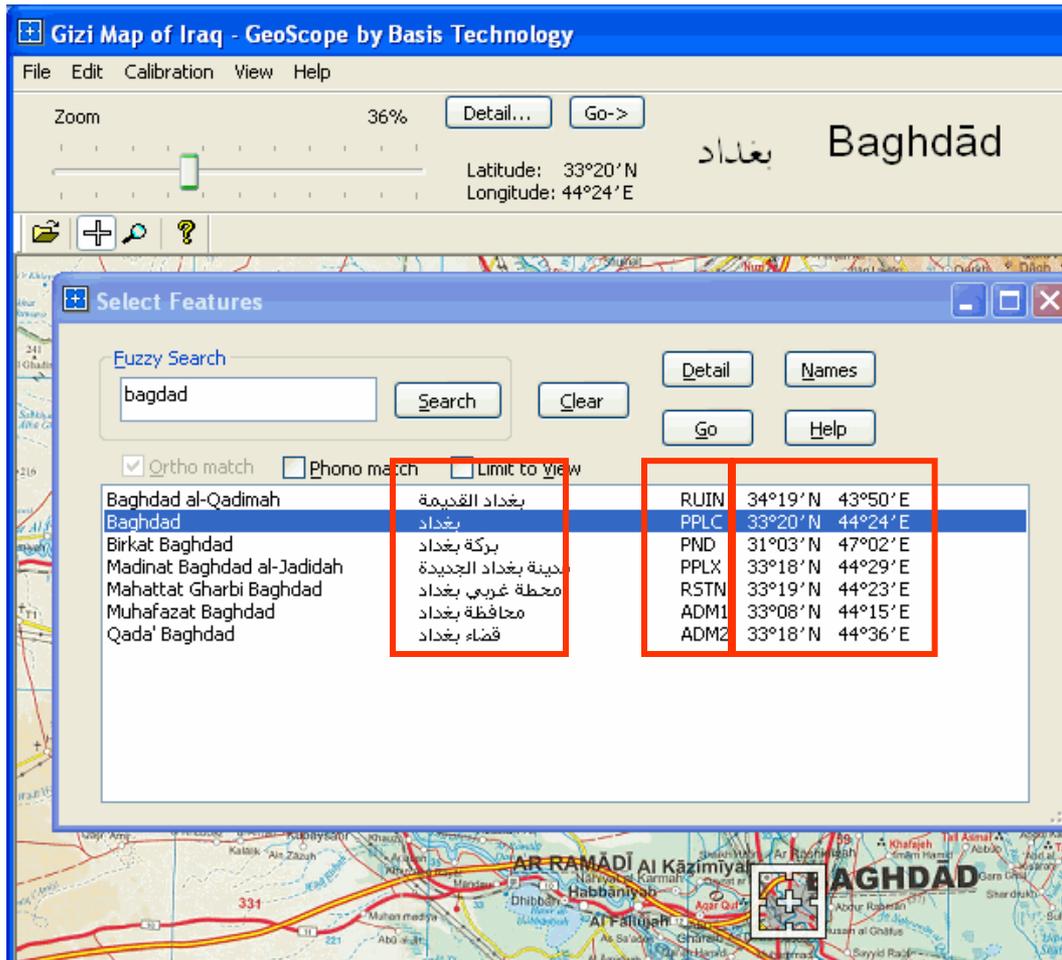
Filter by vowels in input Help

| | |
|-----------|----------------------------|
| السُّفُنُ | 2 parses, 2 vocalizations: |
| السُّفُنُ | (the) ships/vessels |
| as-sufun | NOUN |
| السَّفَنُ | (the) sandpaper |
| as-safan | NOUN |

Ready NUM

- Supports GEOINT analysts focused on the Middle East and Central Asia
- Uses multilingual search technology to locate foreign geographic features
- Includes NGA gazetteers of major geographic features with names rendered authoritatively in native script
- Provides geo-referencing tools for preparing any map

Search Results Dialog



- Arabic name of location/feature
- Type of feature
- Longitude & latitude

- v5.0 scheduled release: October 2012
 - Pashto IC standard
 - Much improved User Interface

- For 2013
 - Mandarin
 - Text Analytics



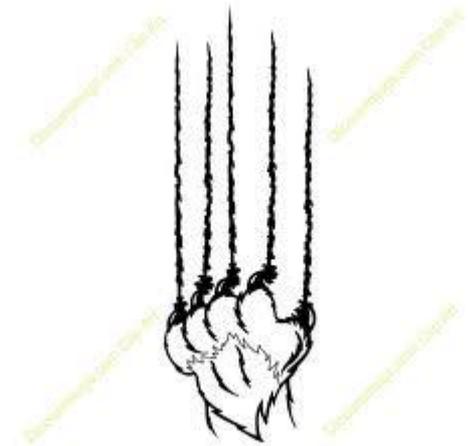
To increase Quality and Production and maximize your limited resources.

5-year program licensed through DIA

Community-wide initiative

Funded by ODNI

- CAT tools improve efficiency and accuracy
 - As much as two-thirds
- ROI of CAT
 - Is different for different agencies hence difficult to formulate





Thank you

For more information:

Visit www.basistech.com

Write to: jflather@basistech.com