Website Transation & Localization

Why you are now better off than your peers…

Congrats!

- You are now have a "value-added" skill…you are no longer <u>just</u> a translator

  - What the *#$%&*!* is he talking about?

Knowing "just" how to translate is *not* enough in today's market...

- As the market for translation increases, so do the number of "translators"

- Globalization (and the Internet) both helps *and* hurts

- Helps:
  - more globalization = more translation

- Hurts:
  - Buyers now see translation as a commodity rather than a "boutique" service
  - Thus, cost has become the primary focal point

DIFFERENTIATION IS KEY…

The average customer is unaware what makes translator "A" different than translator "B "

- How are you different from your peers?

- Do you realize and appreciate those differences?

- Can you explain them to someone else?

Think like an average "Joe" buying your services, not a professional translator…

The Biggest Mistake Translators Make: Playing the "quality" card...

- The *average* translation buyer has no clue about quality
  - Automatically assume (maybe naively) they are getting a quality product

- "*You get what you pay for*" arguments fall on deaf ears

- Degrees, certificates...does anyone even check?
  - When was the last time you asked your plumber to show his license?

- General Rule: Don't sell yourself based on subjective 'facts', sell your skills

Averting the race to the bottom... "value-added" services

- Don't be the cheapest, but back it up with providing services the competition can't
  - Like, well…you guess!

- To succeed, you need to learn (or recognize) skills/specialties **beyond** translation and sell them as services

EXAMPLE
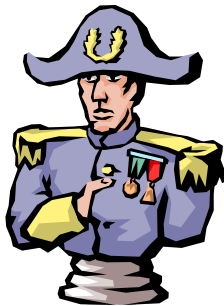
■ Don't just provide translation…find your "value-add"
   ◆ *English > French* Translation Contract for the Foreign Legion

   ◆ French Translator "A" vs…

   = **\$\$\$/€€€ decides winner**

   ◆ French Translator "B"

   ◆ French Translator "A" vs…

   = **Won opportunity for "B"**
   even if* more expensive*

   ◆ French Translator "B" +
   Background in Military Science

   ** especially if you know how to sell yourself

The cold, hard facts...

- The large majority of materials in need of translation in this day and age are authored programmatically or with a software package on a computer.

- Files will be delivered as "softcopy" to the translation provider (individual or agency).

- Your ability to respond to the request, whatever the file format, is essential and reflects on you as a provider of services.
  - Clients don't know a HTML file from a Word document
  - They don't understand when you reject a project even if you explain why
  - Does "I can't work with *[fill in file format here]*" sound like a good business practice to you?
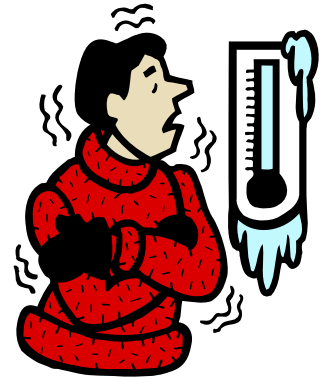
The Problem: Fear, Misunderstanding & Ignorance
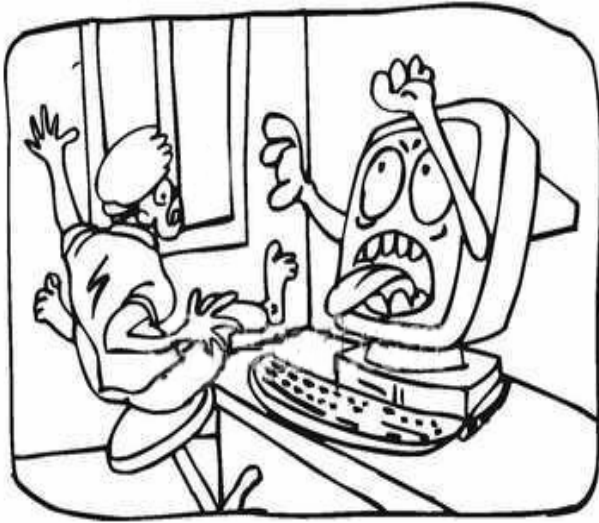
- Your computer IS your closest working companion
  - Spend 8 hours a day working with something and still not understand it? You do it almost every day…

- Majority of file formats (beyond Word & Excel) are largely misunderstood, ignored or rejected by translators

- Ignoring computer "softcopy" formats of all types

  =

  ignoring the face of translation today

Once you understand Web file formats, everything else gets easier

■ Once confidence is built, you can conquer new computer technologies much easier

- ◆ Desktop Publishing applications
- ◆ Advanced Web Formats
- ◆ Software
- ◆ Databases

■ Once you can create basic content in any of these categories, you have the building blocks for translating it!

FACT: Majority of computer software and programming can be learned by just picking up a book or taking a short course

Website Transation & Localization

Client-facing Concepts…

Congrats Again!

- You have essentially graduated from Website Localization Engineering 101…
  - Let's review what we learned:

    - What is Localization (L10N)?

    - What is a "layered" vs. "flat" image?

    - What is an "encoding"?

If you know how to engineer it, you know how to scope it

- Time yourself in advance for an "average" estimate on each task that does *not* involve translation itself:
  - Preparation
  - Post-Processing
  - QA

- I find it best to estimate either by:
  - minutes/file…or…
  - files/hour

- Total hours of work will always vary based on:
  - Complexity of file tree
  - Complexity of authored content
  - Number of files

ALWAYS GET FILES UP FRONT!

- Do NOT fully accept website localization jobs until you have the source in hand!!!
  - It is OK to give rough estimates based on URLs or sample files but…

- Reserve the right to not submit a final price for work until you have files in hand

Ah..those silly (clueless) clients!

"What's a source file?"

"Why do you need to talk to the webmaster? I run the project?"

"Why do you charge for more than just translation?"

"Can't you just type in the text on top of the graphic/picture/image?"

"Can't you just translate the files on the web?"

"Just cut/paste [*or save*] the webpage content into Word and then translate it. I can cut/paste [*or save*] quite easily, why is it a problem for you?"

Always Be Adamant About Being Paid for Work You do BEYOND Translation

- Otherwise you are creating a bad precedent which will be the basis for your relationship

- Calculate your time *against* your word rate and you will realize you are working for pennies

NOTE: Experienced translation agencies *NEVER* accept Website jobs without charging hourly for:
- Engineering time
- Graphic Localization
- *anything* that is labor beyond translation/editing/proofreading (TEP)
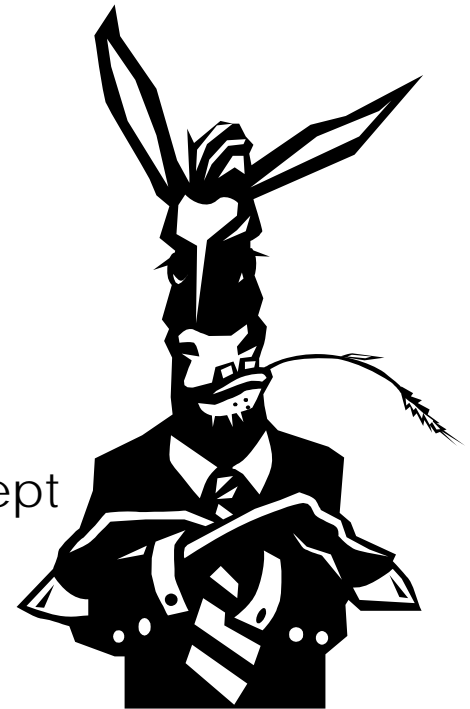
QUALITY, QUALITY, QUALITY...

- Localizing websites is messing with someone's "Picasso"

- If the client is paying for the additional work, you deliver *perfect* quality…period.
  - *perfect* = match the quality of the source

- Visual quality is the most crucial for websites
  - Visual errors are very obvious

- Compare source vs. target on all levels
  - Code
  - Layout
  - Format

Exceptions To the Rules...

- **Web content in format <u>**other than**</u> source**
  - E.g., lines of text in an Excel file or Word
  - Make sure…
    - *they* provide and you are not creating anything
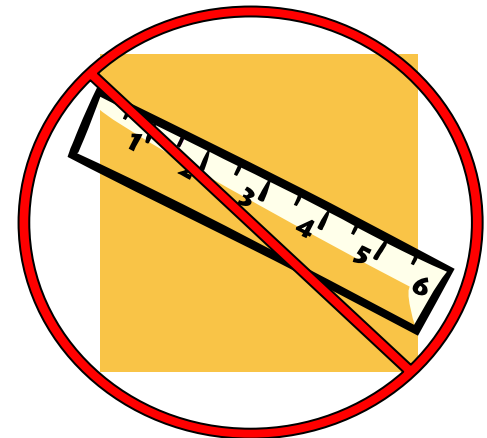   - does not involve any additional labor on your part (like cut/paste)

- **Format provided is "dynamic (server-side)"…**
  - You will recognize them by their file extension:
    - *filename.***asp**   **ASP**
    - *filename.***aspx**   **ASP.NET**
    - *filename.***jsp**   **JSP**
    - *filename.***php**   **PHP**
    - *filename.***cfm**   **ColdFusion**

- **Format provided is XML…**
  - *filename.**xml***   **XML**

Server-Side/ "Dynamic" Websites 101

The Basics…

Quick intro to Dynamic Web Sites

Server sends requested media and text to the browser to be interpreted — 4

User sees web page or external media requested — 5

The user interacts with the web page locally

3. Server accepts request and processes request from browser

6

2. The browser interprets the selection and makes request from appropriate server

1. The client selects a link in a web page

Graphics Text

Data

Sound Video

© Elgendy 2001

- Definition:
  - In it's simplest form, a dynamic website is a website hooked up to a database which both stores and retrieves information and often creates web pages "on the fly."

  - E-commerce (ww.amazon.com), online bank access (www.citibankonline.com), bill pay (www.verizon.com) all are very "dynamic" web sites.

Dynamic Web Sites: Technically

"A web site in which part of the content is generated by server-side code working with a backend database server.

Dynamic pages are not generated until the request comes for a page from the client-side computer.

The request contains parameters, user identity, date and time, context, etc. that are used to create a custom HTML page for those specific parameters."
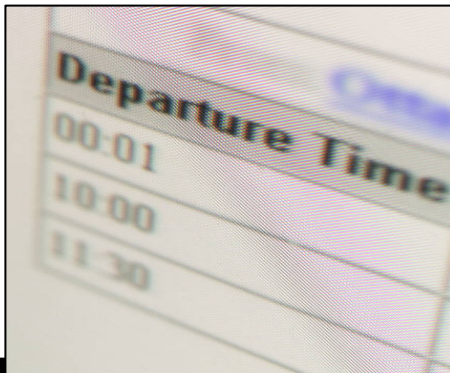
Dynamic Web Sites: Less Technical

The Deli receives your order…

The Server receives your request…

The ingredients are taken from the fridge…

The data is taken from the database…

You provide the "parameters" you want…

You provide the "parameters" you want…

- Sourdough Roll
- Turkey
- Flight from JFK to Cancun
- Lettuce, tomato, pickles, hold the onion
- Dec 26, 2006
- mayo, mustard
- One-way

…and sent to your computer.
…and sent to your door.

…the sandwich is created…
…the webpage is created…

ASP

…the ingredients are put in your sandwich
…the data is pulled into the dynamic server-side page…

Sounds Difficult ☹

- It's not…you know most everything you need to localize "dynamic" server-side pages:
  - Web scripting (like "JavaScript")
    - just beginner knowledge is sufficient
  - ASP(.NET), JSP or PHP
    - just an minor extension of your HTML knowledge
  - Basic Understanding of Databases
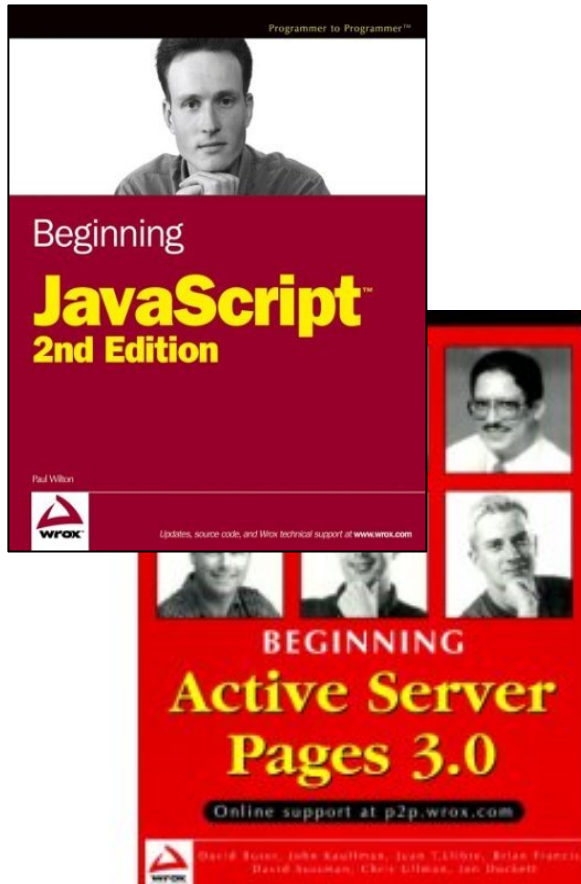   - knowing the basics of Microsoft Access is good enough

Table Of "Dynamic" Formats

Web Format	File Extension	Tag Format
ASP, ASP.NET	.asp, .aspx	<%Tag%>
JSP	.jsp	<%Tag%>
PHP	.php	<?Tag?>
ColdFusion	.cfm	#Tag#
Server Side Include	.shtml, .ssi	<!--Tag-->

■ REMEMBER: Dynamic pages can only be viewed "properly" when running off a server, i.e. they have to be "live."

DYNAMIC WEBSITE REFERENCES

GOOD TO READ:

- Any beginner book on JavaScript or VBScript programming
- Any book on ASP, JSP or PHP

RECOMMENDED:

Any Wrox Press Book…

- **with a high review on Amazon**
- **..and read the intro pages of the book. If it makes sense to you – buy it.**

Highly Recommended

- Beginning JavaScript – *Paul Wilton*
- Beginning Active Server Pages 3.0 – *Many authors*

XML 101

The Basics…

WHAT IS XML?

- A "markup language" along with other "**ML**"s such as HT**ML**
  - Markup languages all look like this:

```
<ContentTag Attribute="Paragraph">

  I am content. I am more content.

</ContentTag>
```

- More flexible than other **M**arkup **L**anguages and can be used for a number of different purposes = e**X**tensible
  - **XML** = e**X**tensible **M**arkup **L**angauge

"XML" is Just Like "A Deck of Cards"

- A deck of cards is a base from which you create "games"

- Predefined general "guidelines" of the deck must be followed:

  - 4 suits (♠♣♥♦)

  - 13 cards per suit
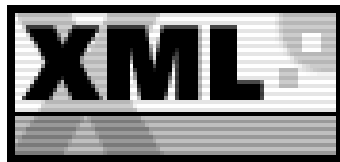
  - 52 cards

  - 2 colors (**red** & **black**)

XML is the Same

- General "guidelines" you must follow to use XML.

  - Non-empty elements are delimited by both a start-tag and an end-tag.

  - Empty elements may be marked with an empty-element (self-closing) tag, such as `<IAmEmpty/>`.
    - This is equal to `<IAmEmpty></IAmEmpty>`.

  - a root element is required

  - end tags are required

  - elements must be properly nested
    - if start tag of A precedes start tag of B, end tag of B must precede end tag of A

  - attribute values must be enclosed in quotes

  - The document complies to its encoding. If no "charset" is defined, usage of UTF-8 is default.

XML...A Deck of Cards?

- These are general guidelines of usage in the specification for XML, but there is no predefined "game" that you are required to play

- XML is just a moldable, formless set of guidelines

XML Specification   Deck of Cards

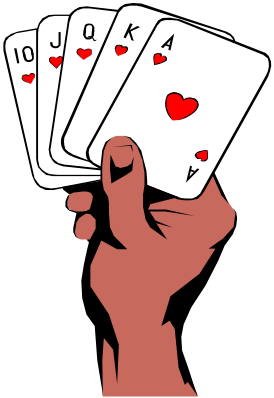| XML Specification	Deck of Cards
■ Tags nested	■ 52 cards
■ Start-tags must have end-tags	■ 4 suits
■ Only One Root Element	■ 13 cards each suit
..etc	..etc.

Creating "Games" from the Deck

- Creating sets of rules around these guidelines allow for millions of "games" from a single deck

- Play an established "game"
  - Like *Poker…*
    - Predictable

- ….or make up your own!
  - A " *House of Cards*"

    - Random

From Poker…

…to a house of cards.

XML is the Same: Well Established "Games"

- The freedom of XML has led to many "games" …

- Many are approved "standards"

Technical Writing
- DITA
- DocBook
- OpenDocument

Software
- RESX

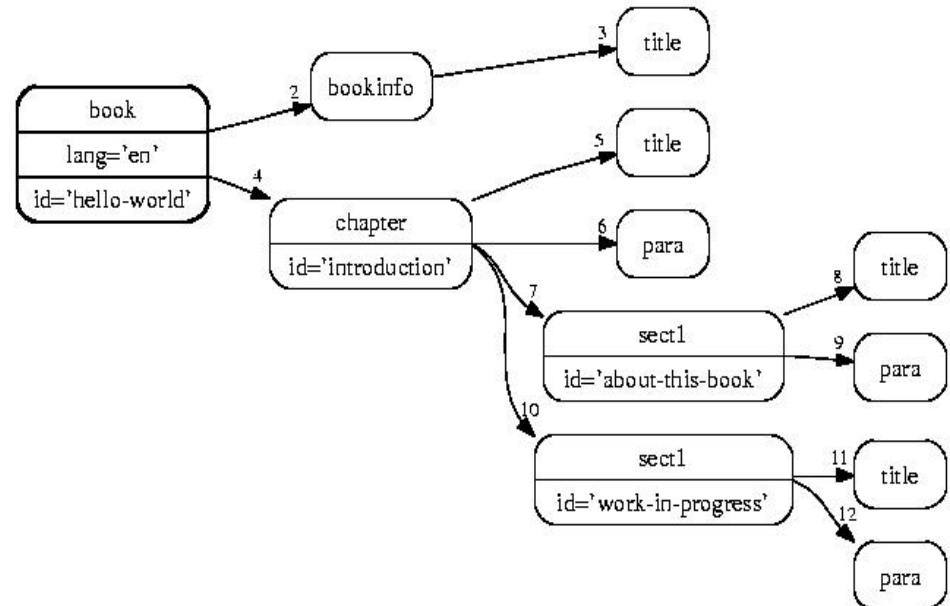*Graphic Design*
- SVG

Translation Industry
- XLIFF
- TMX

DocBook…

XML is the Same: Custom "Games"

- **Others are proprietary based on the application or person that created it**

  - ◆ Adobe InDesign Tagged Format
    - ● INX (**In**Design **X**ML)

  - ◆ SDL Trados
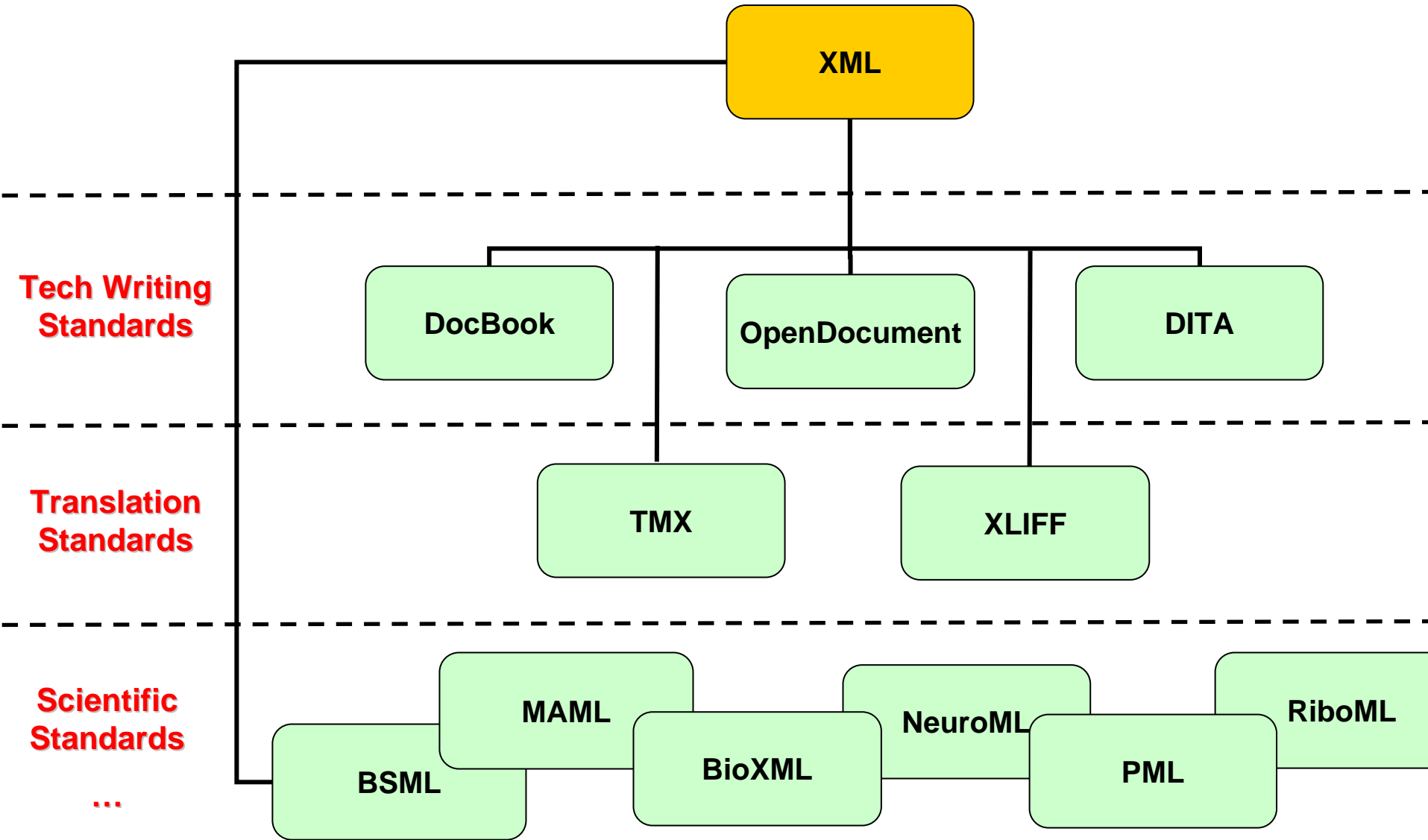   - ● TTX (**T**rados**T**ag **X**ML)

  - ◆ Microsoft Office 2007
    - ● OOXML (**O**ffice **O**pen **XML**)
      - ■ .docx, .xlsx, .pptx

Understanding it All

REVIEW: XML VS. DECK OF CARDS

Saying…   **Is like saying…**
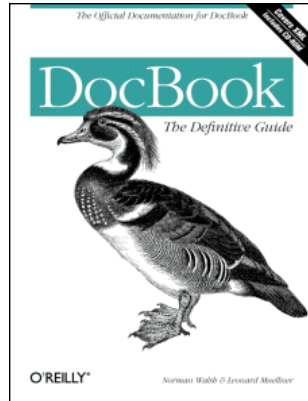
Saying…	Is like saying…
XML	Deck of cards
XML[-based] Standard	Game
"DocBook" (or "DITA", or "SVG", or "TTX")	"Poker" (or "Rummy", or "Bridge")

XML is the Same

- When XML is in compliance with the specification it is called "well-formed"
  - Equivalent to checking that there are 52 cards in your deck and all suits/numbers are present before playing a game
  - Playing with a "full-deck"

Deck of Cards	XML Specification
If…	If…
■ 52 cards	■ Tags nested
■ 4 suits	■ Start-tags must have end-tags
■ 13 cards each suit	■ Only One Root Element
..etc.	..etc.
= "Full Deck"	= "Well Formed"

XML is the Same

- When XML is following the rules of the "standard" or the custom-made rules it is called **"valid"**
  - Equivalent to making sure that no one is cheating
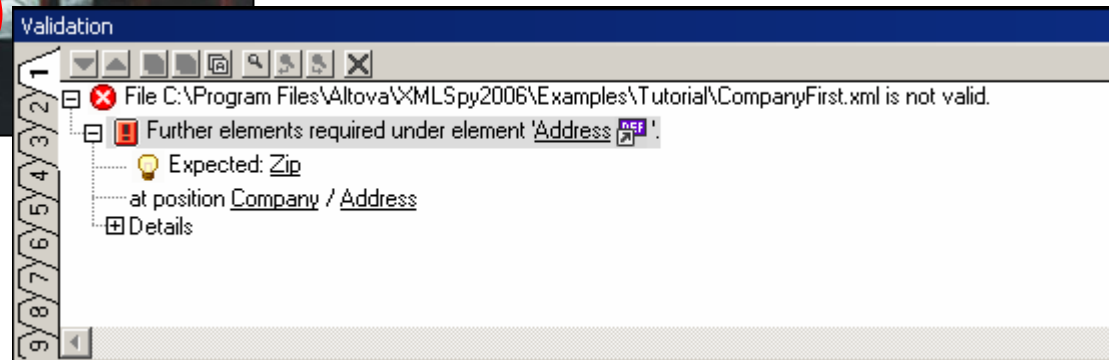
=

Validation

❌ File C:\Program Files\Altova\XMLSpy2006\Examples\Tutorial\CompanyFirst.xml is not valid.
- ⚠ Further elements required under element 'Address'.
  - 💡 Expected: Zip
- at position Company / Address
- ⊞ Details

Content Management Systems

Getting a Grip

What is CMS/GMS

- Tell me what you think a CMS is…
  - You're all right! There really is no hard and fast definition.

- Basically ALL just big, enterprise-level databases that can centralize and store content, be it:
  - Documentation
  - Graphics
  - Web Content

- A very broad, all-inclusive category of products
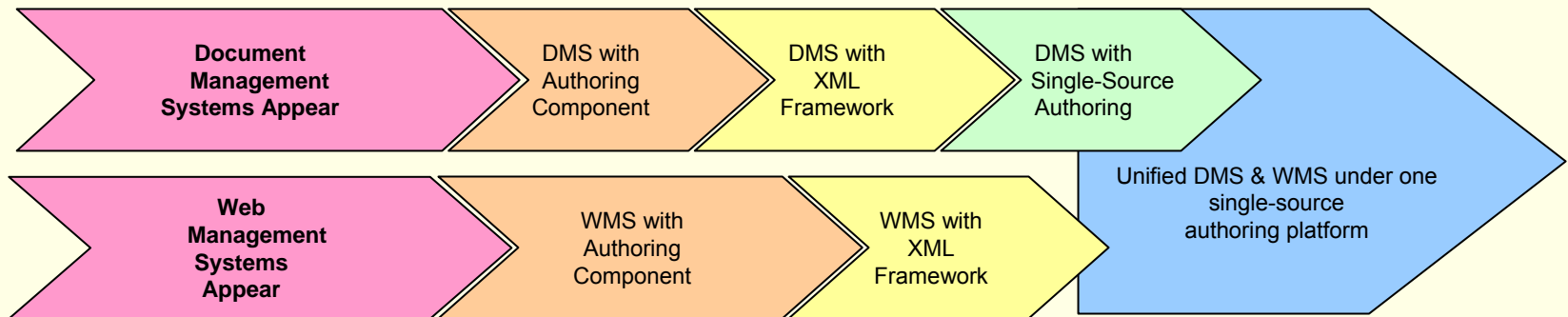
Best Definition I Have Found

- A content management system (CMS) is a data-storage system with several functions which allows users to:
    - Collect, maintain and publish information
    - to manage content lifecycles
        - Versions
        - Archives
    - to manage access to content by appropriate users
    - and to route tasks such as content creation or approval among users as appropriate.

CMS: Evolution

CMS Evolutionary Stages (From A L10N Perspective)

Document Management Systems Appear → DMS with Authoring Component → DMS with XML Framework → DMS with Single-Source Authoring

Web Management Systems Appear → WMS with Authoring Component → WMS with XML Framework

Unified DMS & WMS under one single-source authoring platform

Mid 90's   **Late 90's**

Early 00's   **Mid 00's**

Future (Late 00's?)

CMS *Primarily* Export XML

- XML is the most common export format for most CMS on the market

...BUT XML is Rarely The Final Output Required.

- XML is really just an intermediary file format

- The XML is a temporary container for the content
  - Like the cocoon of a butterfly
  - a "shell" from which more beautiful "creatures" are created

- It is ultimately used to create more robust outputs:
  - a printed manual
  - a help system
  - a webpage
  - another file format
  - ...anything the customer has their CMS set up for
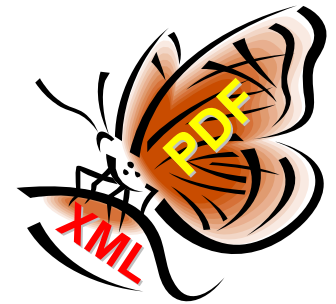
Website Transation & Localization

Applications of Interest…

Key: $ = $50 or less
                                     $$ = up to $200
                                   $$$ = more than $200

Word Counting/ Quoting Project Scope
Used: N/A

- Use your CAT tool of choice for analysis
  - SDL Trados
  - SDLX
  - DejaVu
  - Transit…etc.
- Other "specialized" tools:
  - **Free**
    - FineCount
    - Translator's Abacus
  - **$**
    - AnyCount
    - PractiCount
  - **$$$**
    - WebBudget

Web Preparation/Translation
Used: "ENLASO Rainbow"

- Use your CAT tool of choice for preparation
  - SDL Trados
  - SDLX
  - DejaVu
  - Transit…etc.

- Other "specialized" tools:
  - **Free**
    - Cat's Cradle
    - **Rainbow**
  - **$$$**
    - WebBudget
    - Catalyst (latest version)

Graphics Preparation
Used: "XnView"
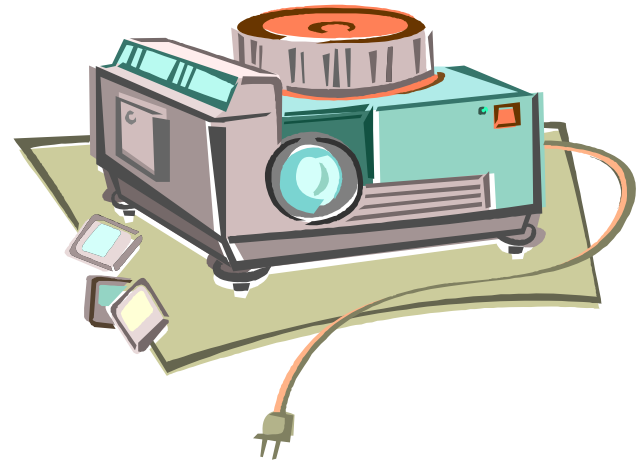
- Image browsing/viewing software packages:
  - **Free**
    - XnView
    - FastStone
    - IrfanView
  - **$**
    - ACDSee
  - **$$**
    - Photoshop *"Elements"*

Graphics Localization
Used: "Photoshop"

- Image editing software packages:
  - **Free**
    - Gimp
    - Paint.NET

      WARNING: All of the free applications do NOT allow for tight control of text layers (kerning, leading, width, etc.)

  - **$**
    - Paint Shop Pro

      A good, cheaper PhotoShop alternative

  - **$$$**
    - Photoshop (the "real thing" NOT "Elements")

      Used by 98+% of graphic professionals

WEB CONTENT CODE QA
USED: "TAGCHECK & XENU LINK SLEUTH"

- Other "specialized" tools:
  - Free
    - TagCheck + Xenu
    - Latest Version of Rainbow (Okapi Rainbow)

    Can create your own custom QA scripts to follow.
  - $
    - BeyondCompare

    A source code comparison tool. Can create your own custom QA scripts to follow.
  - $$$
    - HTML QA

Web Content Visual QA
Used: "Horizon"

- Other "specialized" tools:
  - ◆ Free
    - ● Horizon
  - ◆ $
    - ● BeyondCompare
    - **You can also see web pages loaded side-by-side**