How to help search engines indexing photos

Michael Steidl, IPTC
Photo Metadata Conference 2012
Session 1
17 May 2012, London

What we mean by “searching”

We type in search terms (= text) and see a list of results
Simple Rules

• Search engines index primarily text of content found on the web
• Even if you want to find a photo you use a term for searching = a text string
• A search engine shows as search results
  – Text content: text on web pages, PDFs, Word documents etc.
  – Visual content: photos, videos

Big question

How does a search engine find pictures if one searches for text?
Answer

- Search engines *do not* read metadata embedded into image files – that's reality.
- They assume that text next to an image on a web page tells something about the content of the image.
- This assumption can be improved by semantic mark-up

How to tackle this?

- Cry and squeeze the IPTC stressball
- Or take action → see the examples
Example 1

- UK newspaper 1
- Plain HTML mark-up
- Plain search result display

Example 2

- UK newspaper
- HTML mark-up
- + Semantic Mark-up of schema.org
- Enhanced display of search results
Example 3

Behind Twists of Diplomacy in the Case of a Chinese Dissident

- US newspaper
- HTML mark-up
- + Semantic mark-up using schema.org + IPTC’s rNews

What makes the difference

- Plain HTML only page:
- Searching for a text which is “not” related to a photo
- … but you find this photo
What makes the difference 2a

• Semantic mark-up page:
• Searching for a text which is *not* related to a photo
• ... and you do not find this photo – only related ones

Behind Twists of Diplomacy in the Case of a Chinese Dissident

Days of drama preceded the agreement between the United States and China in the case of Chen Guangcheng, but the negotiations ultimately reflected a mutually beneficial relationship, officials said.

Indians Host Clinton While Also Wooing Iran

This week’s diplomatic choreography — with Secretary of State Hillary Rodham Clinton and the Iranian's new leader — provided a window into the evolving relationship between India and Iran.

Activist Says Chinese Authorities Are Assisting His Plans to Go to U.S.

Chen Guangcheng, whose flight to the American Embassy triggered a diplomatic crisis, told Time that Chinese authorities have begun to assist him in applying for permission to travel to the United States.

What makes the difference 2b

• Semantic mark-up page:
• Searching for a text which *is* related to a photo
• ... and you find this photo

Behind Twists of Diplomacy in the Case of a Chinese Dissident

Days of drama preceded the agreement between the United States and China in the case of Chen Guangcheng, but the negotiations ultimately reflected a mutually beneficial relationship, officials said.

Indians Host Clinton While Also Wooing Iran

This week’s diplomatic choreography — with Secretary of State Hillary Rodham Clinton and the Iranian's new leader — provided a window into the evolving relationship between India and Iran.

Activist Says Chinese Authorities Are Assisting His Plans to Go to U.S.

Chen Guangcheng, whose flight to the American Embassy triggered a diplomatic crisis, told Time that Chinese authorities have begun to assist him in applying for permission to travel to the United States.
Conclusion

- Web pages with plain HTML mark-up establish a lightweight relationship between text and photos on this page.
- Semantic mark-up can establish a strong and unambiguous relationship between text – and text portions – and photos.
- Goal:
  - Add photo metadata to a web page, e.g. the caption.
  - Add semantic mark-up to indicate e.g. "this is the caption of the photo".
  - Add semantic mark-up to indicate: this photo relates to that article.

How can this be easily achieved?

- Join the Breakout Session "What will image search look like in the future?"
- and check the embeddedmetadata.org website.