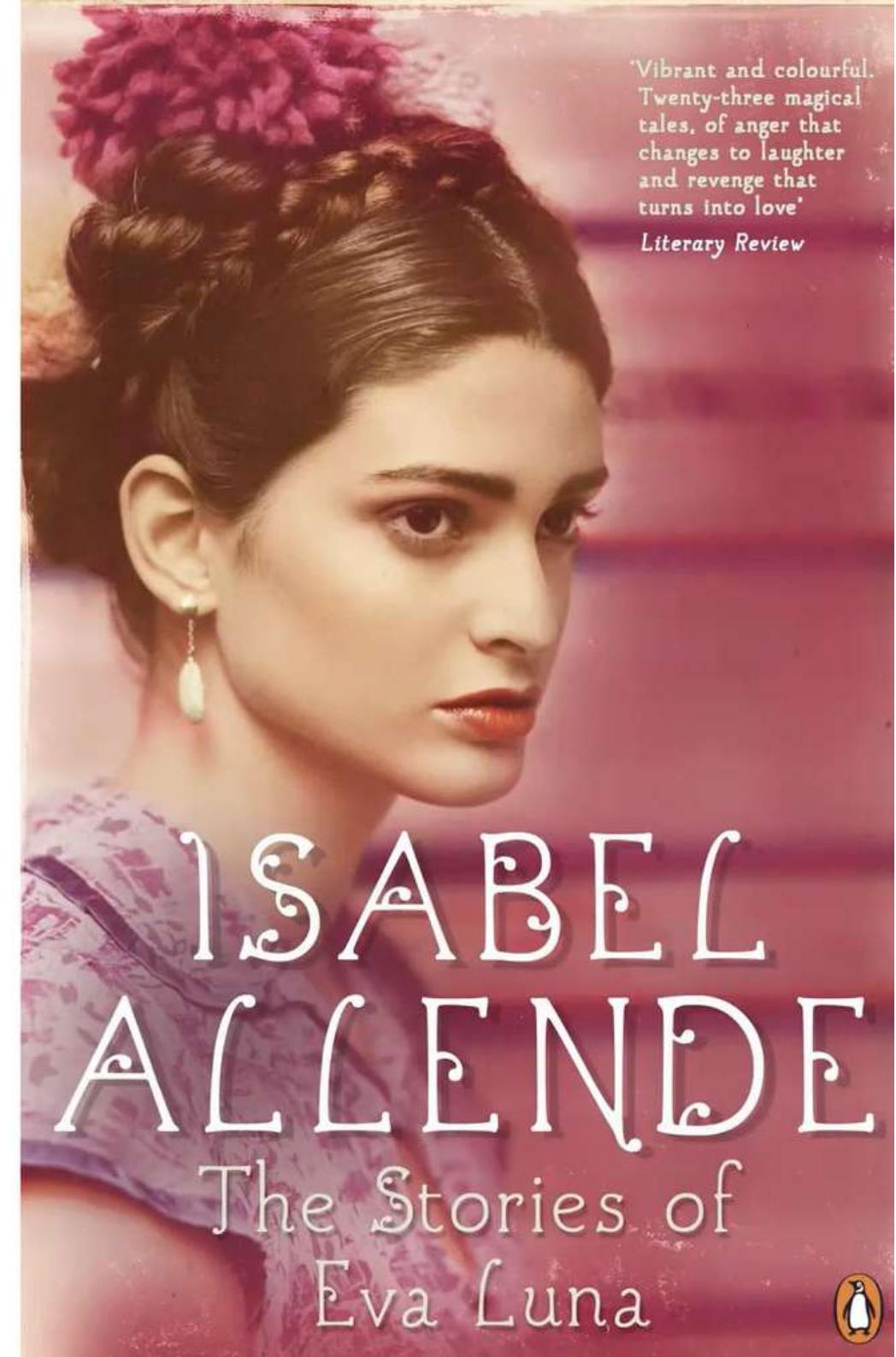


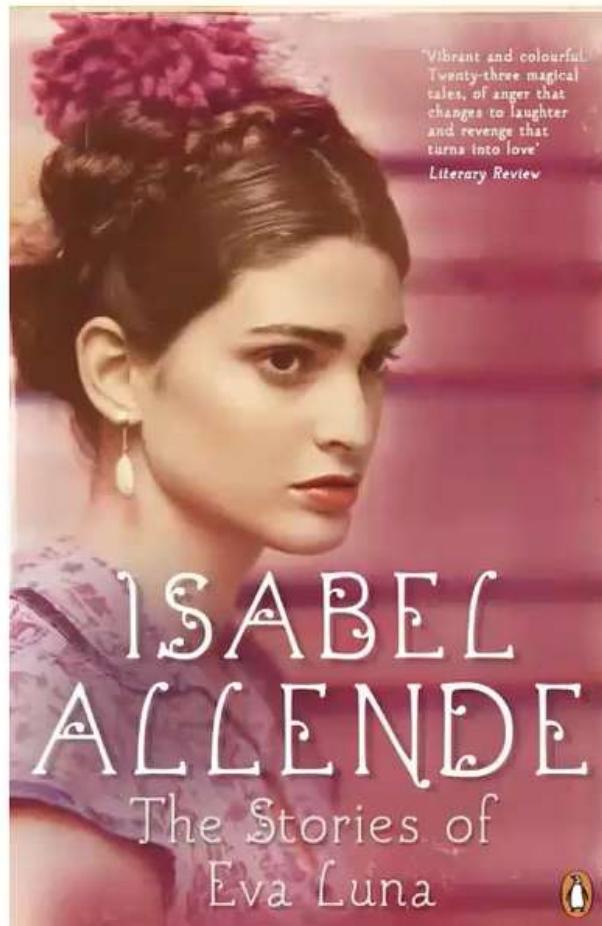
# Names

- ◆ “Two words” tale  
(read it if you can)



# Names

- ◆ Let's talk about names on the web, both sides: *social* and *technical*



# Web names: social side

- ◆ How to choose a ***good web address*** ?
- ◆ There are some rules that can maximize the potential success of the site
- ◆ How much? (read as: shoud we care about a name?? Can it really impact a website??)
- ◆ Answer: on average **10-20%**, with variance up to **50%** (!!)

# Rule 1

- ◆ Shorter names work much better than longer names

## Rule 2

- ◆ The name should be unique, sufficiently separatable from other names (that is: it shuld not be confounded with other names!)
- ◆ For instance: never choose a *plural* when the singular is already taken!!

# Rule 3

- ◆ Take the .COM (!!!!)
- ◆ Impact: about **+4.5%**

# Rule 4

- ◆ It should be easy to memorize and write

# Rule 5

- ◆ (overlapping with rule n.4): better to choose existing words rather than creating new ones
- ◆ In any case, if new words, or acronyms, are used, the important factor is the ratio between the «standard» words and the new ones: the higher the better
- ◆ Range: from +1.5% to -5% (!)

# Rule 6

- ◆ Beware the sound...!
- ◆ Intuitive rule: it must «sound well»
- ◆ In practice?

# The beginning of a name....

- ◆ Names that start with a wovel work well  
(about **+3.7%**)
- ◆ Semi-wovel (r, j, y, w) work well  
(about +2.9%)
- ◆ Surprise: consonants like f, v, s, z work even better (**+3.3%**)
- ◆ Surprise 2: the consonants p, k, t work better than the remaining ones (+1.9%)

# Sound (cont.)...

- ◆ Sounds associated with bad words in English (like the phonem «uh», like in «yuck») damage the site up to a – **44% (!)**
- ◆ In other contexts though, for instance adult material (!), they give advantages up to **+7% (!!)**

# Rule 7

- ◆ No dashes («-») !!!
- ◆ Impact: about **-3%** !

# Ruke 8

- ◆ No numbers  
(rule that is also present in the Internet)
- ◆ **FALSE!**
- ◆ In fact: use numbers!  
(Impact: about **+8.2%** !!!!)

# Rule 0

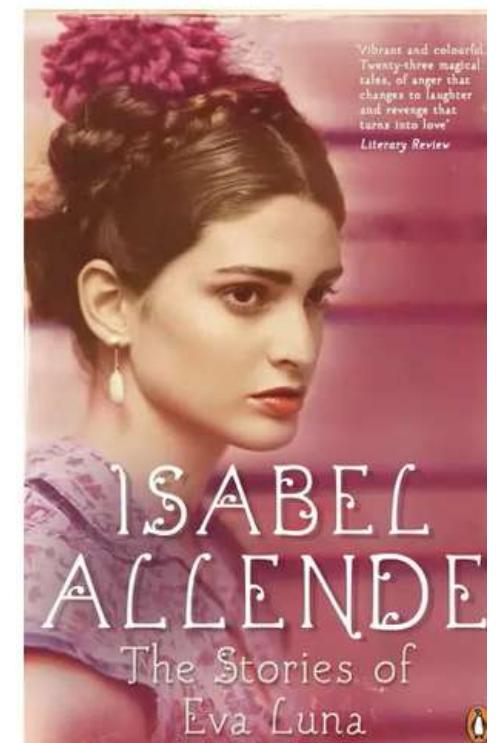
- ◆ Be careful on how you check whether a domain is free or not... (!!!!!!!!)
- ◆ Suggestion: **internic**

# Web names: technical side

- ◆ **BUT WE ALREADY KNOW  
EVERYTHING ABOUT WEB  
ADDRESSES....!!!**
- ◆ **http www ilikeyou dot com!**
- ◆ They are... that thing... how is it called  
technically... ah yes, URL ☺
- ◆ Sure that's enough?

# URI!!

- ◆ Uniform Resource Identifier
- ◆ The real milestone, the “names” that identify the “things” (resources)



# URIs... a new thing?

- ◆ Defined in **RFC2396** (1998), RFC3986 (2005)
- ◆ ... impressing the number of people working on the Web that don't know about URIs...

# URI and Web addresses

- ◆ URI == web addresses http://... ?
- ◆ **NO!**
- ◆ URIs are a *superset* of the web addresses http://....

# Web addresses

- ◆ Common use for instance to say  
`www.mysite.com`
- ◆ This is not a URI!
- ◆ And in fact... the form above doesn't exist in  
any standard (!)
- ◆ It's the browsers that, for convenience,  
translate this "web address" (for the people)  
into `http://www.mysite.com/`
- ◆ ... which is a URI

# Examples of URIs

- ◆ <http://corsi.math.unipd.it/wim>
- ◆ [news:it.culture](#)
- ◆ <telnet://example.net:453>
- ◆ <mailto:massimo@gmail.com>

# URIs, URLs, URNs

- ◆ URIs are the most general Web identifiers
- ◆ Then, some URIs can be **URLs** and/or **URNs**

# URLs

- ◆ Uniform Resource **LOCATOR**
- ◆ URLs are those URI that identify the resource via a representation of their primary ***access mechanism***
- ◆ So, yes, for instance web addresses like “`http://....`” are URLs

# URNs

- ◆ Uniform Resource **NAMES**
- ◆ URNs are those URIs that stay unique and persistent even when the resource doesn't exist any more or it not available any more

# URI structure (absolute form)

◆ **Schema : part-depending-on-the-schema**

# The Schema

- ◆ Defines the semantics (the meaning) of the URI
- ◆ For instance
- ◆ In <http://corsi.math.unipd.it/wim>
- ◆ ... the schema is **http**
- ◆ → This is a URI whose semantics is set by the formal specification describing the http schema

# URI (cont.)

- ◆ URIs can be:
  - ◆ *hierarchical*
  - ◆ *opaque*

# Hierarchical URIs

- ◆ Typical general form:
- ◆ **Schema :// authority path ? query**

# Authority

- ◆ Literally, the authority!!
- ◆ The element that states that URI is under control of a certain authority



# Example

- ◆ <http://corsi.math.unipd.it/wim>
- ◆ http → schema
- ◆ corsi.math.unipd.it → authority



# Path

- ◆ Again, literally, the path
- ◆ Composed of zero or more *segments*,  
each of the form
- ◆ **/segment**



# Example



- ◆ [ftp://library.site.com/books/Ken\\_Follett/ThePillarsOfTheEarth.txt](ftp://library.site.com/books/Ken_Follett/ThePillarsOfTheEarth.txt)
- ◆ ftp library.site.com → schema
- /books → authority
- /Ken\_Follett → segment
- /ThePillarsOfTheEarth.txt → segment

# Note: the hash(#)

- ◆ “#” is a reserved characters in URIs
- ◆ Needed to separate the URI of an object with an identifier to a fragment of the object: so URIs can refer not only to a resource, but to a subpart too.

# Query

- ◆ Information interpreted by the resource (“input parameters”)



# Example



- ◆ <http://www.mytravel.com/timetable?start=Padua&destination=Rome>
- ◆ http → schema  
www.mytravel.com → authority  
/timetable → segment  
start=Padua&destination=Rome → query

# Opaque URI

- ◆ Typical general form:
- ◆ **Schema : opaque\_part**



# Opaque part?

- ◆ Roughly like the path of the hierarchical URI, just without “/”



# Example

- ◆ mailto:director@cnn.com
- ◆ mailto → schema
- ◆ director@cnn.com → opaque\_part



# More URIs

- ◆ **TELEPHONE** (/FAX etc etc)
- ◆ Rfc2806
- ◆ Examples:
- ◆ **tel:+358-555-1234567**
- ◆ **fax:+358.555.1234567**
- ◆ **modem:+3585551234567;type=v32b?7e1;type=v110**

# URI (cont.)



- ◆ A URI can be **absolute** or **relative**
- ◆ Absolute: already “complete” as it is
- ◆ Relative: incomplete. To be completed it has to be turned into an absolute URI via information deriving from the ***context***

# Beware...



- ◆ To use relative URIs that are completed into something we don't want
- ◆ Example of one of the most common web errors: in our `http://www.mysite.com` we insert a link to “`www.othersite.com`”
- ◆ → completed into  
`“http:www.mysite.com/www.othersite.com”`

# Example (cont.)...



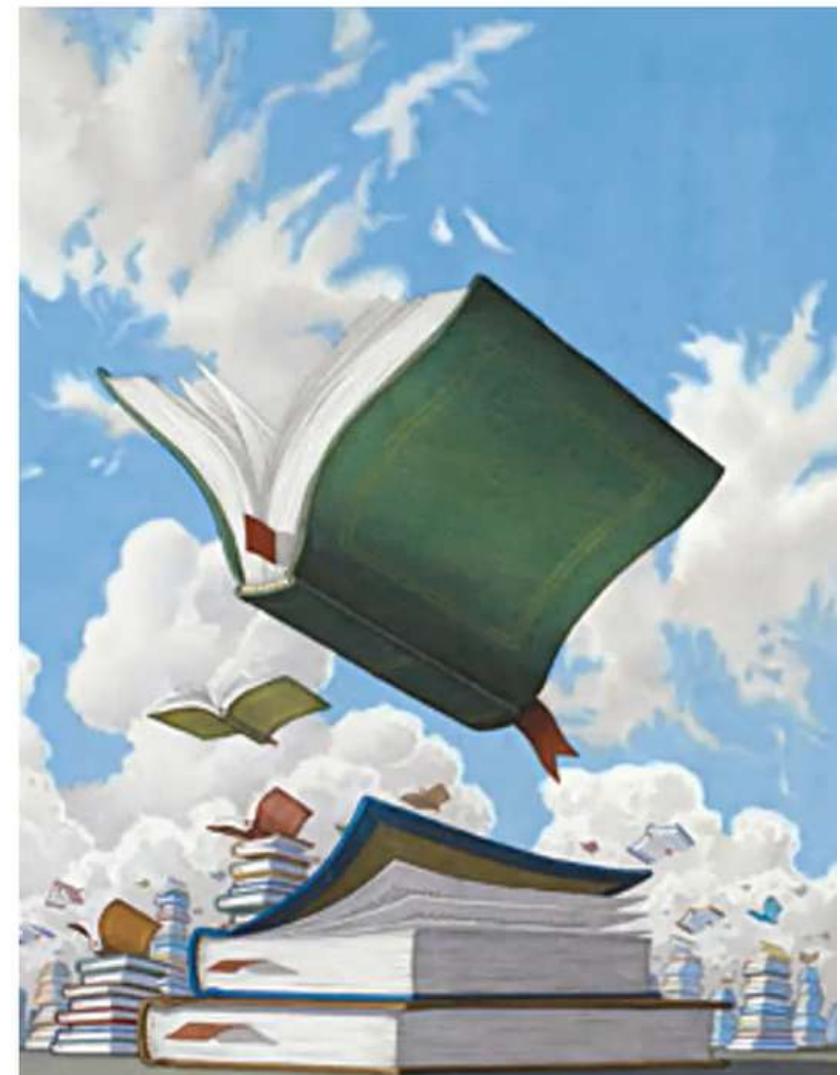
- ◆ Within `http://www.mysite.com`, we insert in a web page the email link to `"shop@mysite.com"`
- ◆ → completed into  
`"http://www.mysite.com/shop@mysite.com"`

# URNs

- ◆ Uniform Resource **NAMES**
- ◆ Typical form:  
**urn:NID :....**
- ◆ Where *NID* is the so-called *Namespace Identifier* : in a sense the «schema» of the URN

# Example: ISBN codes

- ◆ 0-395-36341-1
- ◆ Group-of-editors
- ◆ Editor identifier
- ◆ Title identifier
- ◆ Checkdigit (mod.11)



# ISBN as URNs

- ◆ Rfc3187
- ◆ URN:ISBN:0-395-36341-1



# New Zealand...

- ◆ Has been the first state to ask and obtain a URN of its own (!)
- ◆ urn:nzl:...



# Beyond URI

- ◆ URI: URLs URNs etc
- ◆ But.... what **encoding?**
- ◆ ASCII...(!) ☹
- ◆ → **IRI**  
(Internationalized  
Resource Identifiers)

實 神 活 愛 家 音 上

# IRI and security...

- ◆ IDN (Internationalized Domain Names), even in extensions: example .рф (.ru)
- ◆ ***Homograph attack***  
(rnicrosoft.com, G00GLE.C0M...)
- ◆ → enormously powered up with IRIIs  
(es. Cyrillic, Greek etc)

wikipedia.org

# Back to URI: problem

- ◆ What is the data format of  
<http://www.sito.it/a/b.html>?  
(Italian? English? French?)



# Hmmm...

- ◆ Think of this situation:
- ◆ "Queen Elizabeth Street"
- ◆ → everybody living there  
is a noble??



# The Answer

- ◆ That information is not part of the URL/URI: the web address is a black-box string, and every property depends on the schema, not on the string (!)
- ◆ HTTP provide methods (***content negotiation***) to transmite the correct data format. Every other algorithm (like using the final extension of the URL, like .txt, .gif etc) is not 100% reliable dell'URL non è attendibile.

# URI problems: human idiocy

- ◆ ... via the TLDs
- ◆ **TLD** =  
**Top Level Domain**  
(like .com , .net, .org)

# “TLD overkill”

- ◆ Create a new TLD to solve a problem

# Example of problem:

- ◆ In the Internet it's full of crazy perverts, lots of sex and vulgar people!!
- ◆ Such a bad place, compared for instance to tv!



In fact, look at what's in the Internet...









X

...

VH1 #DATINGNAKED

THURS JULY 17 + 9/8C





**.XXX**

- ◆ So: proposal to add a new TLD:  
**.xxx** = “pornographic” site
- ◆ ( or, **.sex**)



# And dually...

- ◆ Proposal to add a **.kids** ("secure" site)



# Other proposal

- ◆ “.adult”
- ◆ → more generic, “adult website”



# Proposal... by whom?

- ◆ ICM Registry, California-based company  
(Palm Beach Gardens)



# So we can implement this two ways...

- ◆ ["light"] Adding domains with these new TLDs, or
- ◆ ["strong"] Force all the content of a specific kind into the corresponding TLD



# Costs?

- ◆ Technological cost: **SUPER LOW!**
- ◆ Social cost: ... (!?)

# Pro/cons ("light" case)

- ◆ **Pro:** ???
- ◆ **Con:** companies will (maybe) buy another .sex domain, keeping the existing one
- ◆ *Question:* But that's just a transition, eventually they will all migrate, no?
- ◆ *Answer:* no! (tracking/filtering)

# Pro/cons (“strong” case)

- ◆ **Pro:** the WHO axis now gets lot of extra information by the TLD
- ◆ **~~Pro:** no more problems with sex/vulgarity & co. on the Internet...
- ◆ .....
- ◆ .....
- ◆ ..... **(YEA SURE!!)**

# Problem....

- ◆ For instance, how do you define what is “porno”?

# Example...



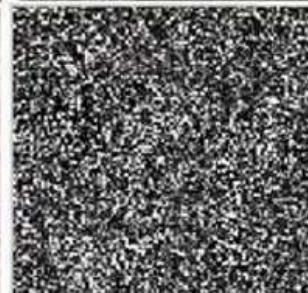
# Another example...



# What is this?



- ◆ The most famous test image used in computer graphics (great mixture of details, “flat” regions, shading, textures)
- ◆ Time ago it even came included in every graphic program (like for instance within Windows with Paint)



# The lady?

◆ Lena Soderberg (ne Sjööblom),  
Swedish, happily living there, married  
with three children, working in the state  
monopoly and helping as a volunteer  
for disabled people



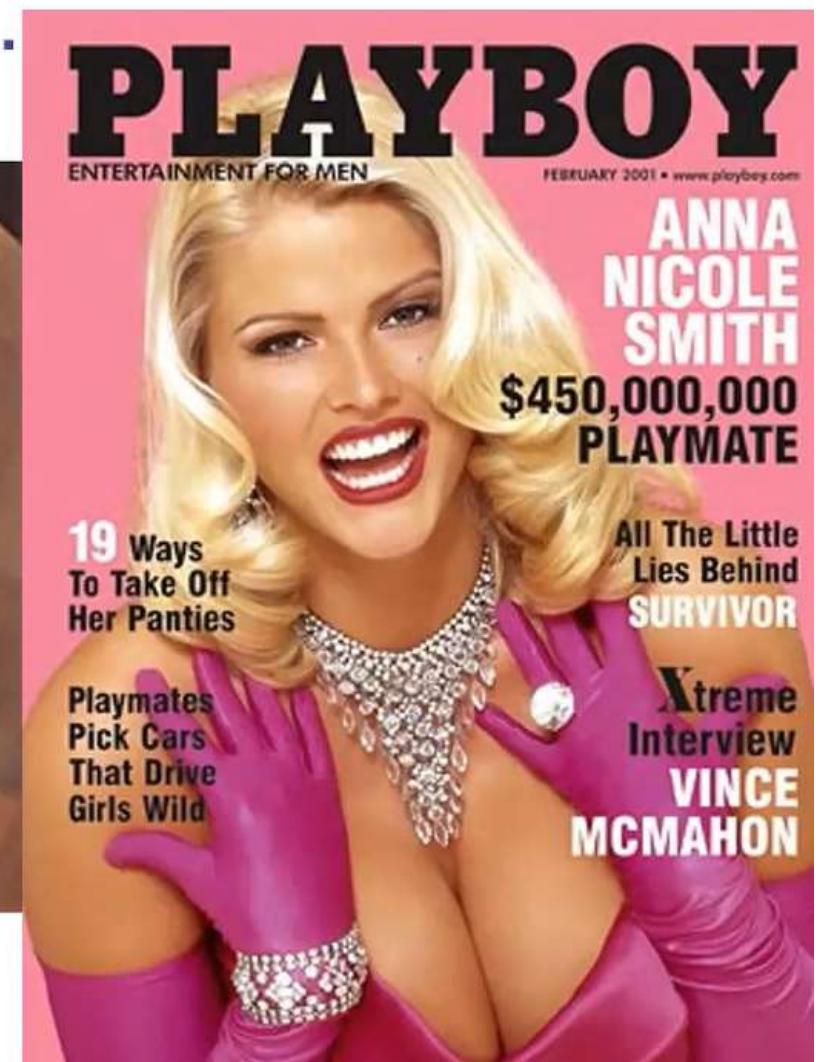


# Back to that image

- ◆ Scanned by programmers of the University of Southern California, to first use it as test image for digital compression and transmission tests on Arpanet (Internet's predecessor)
- ◆ Scanned from where?

# Playboy!

- ◆ ... scanned from the big centerfold page of Playboy (1972!!) and ...



... “cropped”...



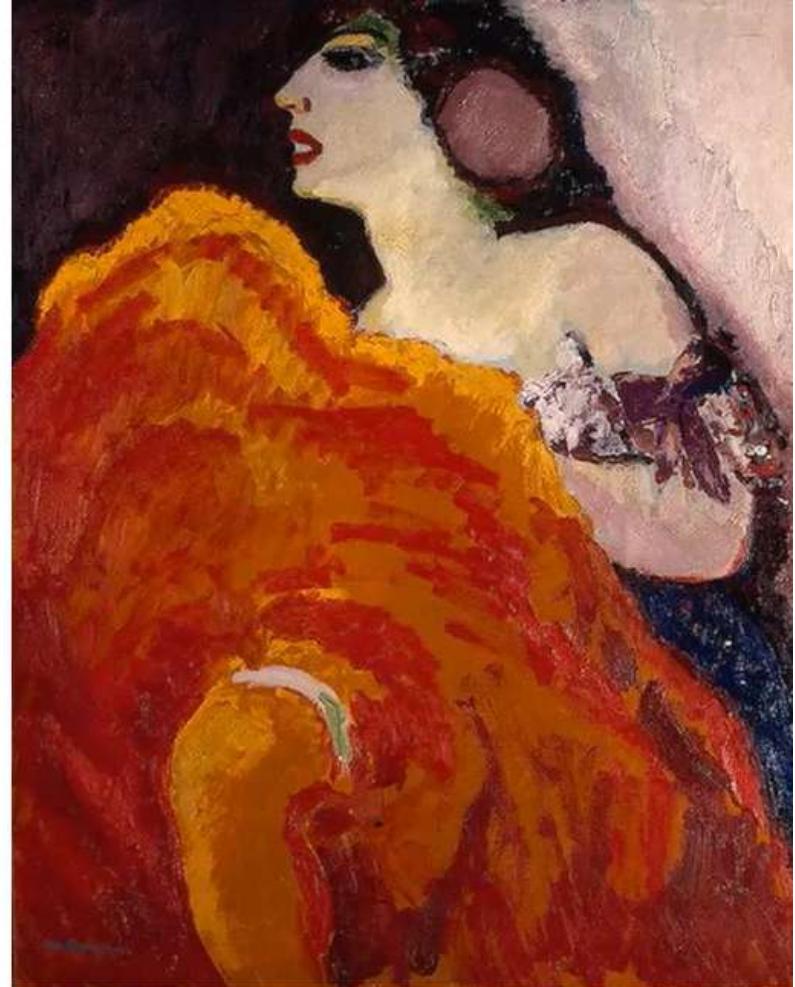
# And so...



- ◆ In 1996 someone curious finds out
- ◆ → petition to ban forever that (for instance from all articles in the  
**IEEE TRANSACTIONS ON IMAGE PROCESSING**)
- ◆ Formally never “forbidden”, but “strongly suggested” not to use it → people started to use it less and less → taken out from all graphic programs
- ◆ → ***social pressure*** can radically change perception of a resource along time (!)

# Boundaries...

- ◆ So what is porno and what is not? (Renoir, Picasso, Dali', Botticelli, a Fellini movie?)





≡ MENU Q CERCA

IL QUOTIDIANO S ABBONATI

Sei qui: Home &gt; Cronaca

# Facebook "censura" la pagina del Museo di Ravenna per un nudo artistico

La denuncia dell'ente: «Uno scatto di Paolo Roversi, un'opera d'arte di grande valore, è stato identificato e riconosciuto come contenuto pornografico»

24 Giugno 2021 | Modificato il: 24 Giugno 2021 | 1 minuti di lettura



**D**a circa un mese Facebook ha oscurato la pagina del Museo d'arte della città di Ravenna dopo che era stata pubblicata un'immagine di nudo artistico esposta durante la mostra "Paolo Roversi.

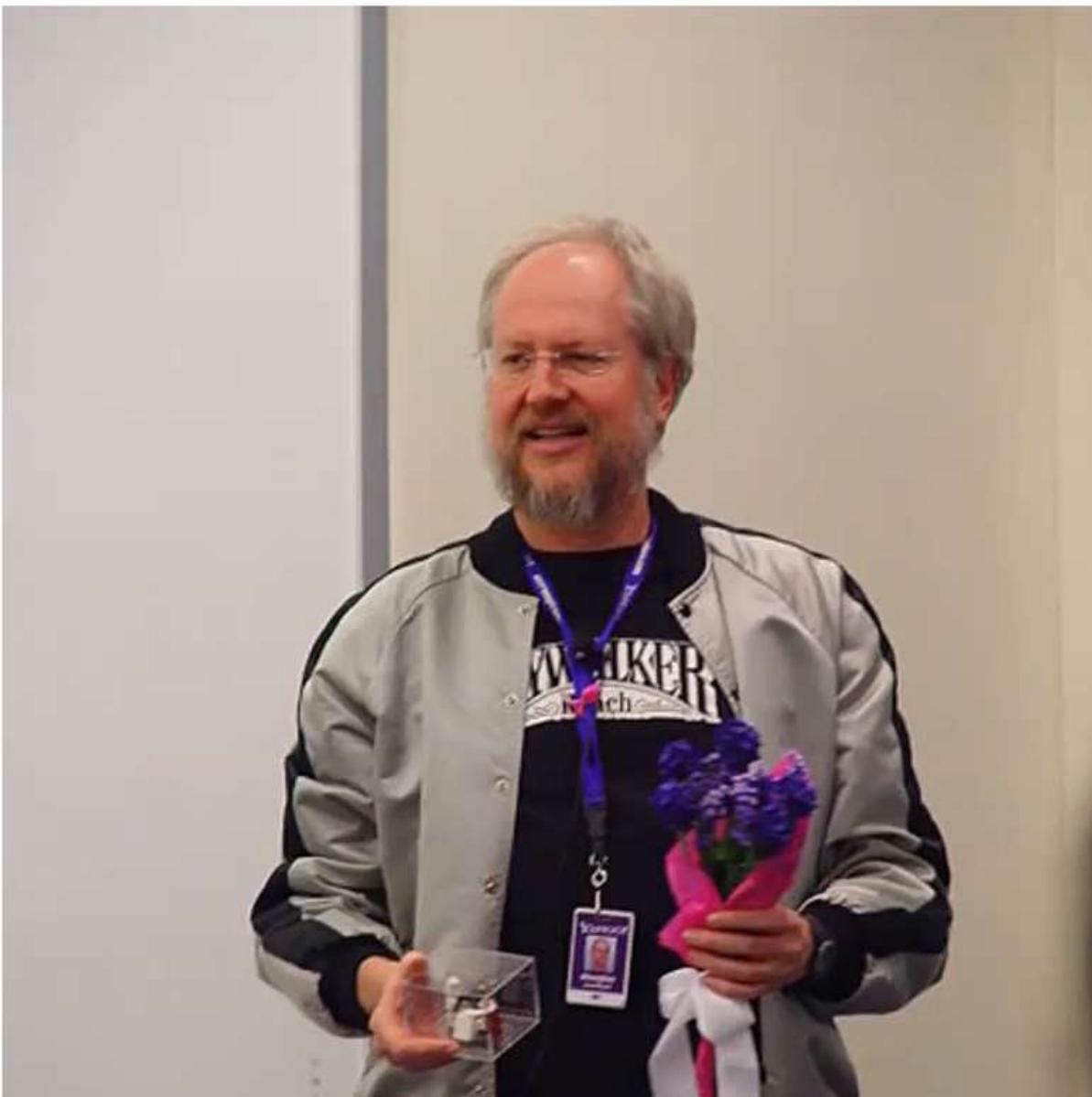
"Studio Luce" conclusasi il 6 giugno scorso. Uno scatto «identificato e riconosciuto - denuncia il museo in una nota - dall'algoritmo di Facebook come contenuto pornografico. Peccato che nella realtà si trattasse di uno scatto d'autore firmato dal grande fotografo ravennate Paolo Roversi, un'opera d'arte di grande valore. Verrebbe anche da passarci sopra, visto che la censura nasce da un algoritmo meccanico e non da un

VIDEO DEL GIORNO



Circostanzio: Gualtieri sistema le buche stradali, ora bisogna convincere i romani che sta accadendo davvero

# Related problem on “who decides”...



# Douglas Crockford

- ◆ Co-inventor of Javascript, JSON etc... (!)



# JSMIn

- ◆ The reference library to minimize Javascript
- ◆ License, free software with an added sentence:
- ◆ “The Software shall be used for Good, not Evil”



# Result

- ◆ Infinite legal battles and argumentations... (!), whole operating systems incorporating that library at risk... (!!!)



# SQLite

◆ “May you do good and not evil  
May you find forgiveness for yourself  
and forgive others  
May you share freely, never taking more  
than you give.”



# Back to URL: the moral

- ◆ Using URLs for scopes that go beyond, forcing information that doesn't belong there, is just stupid!



# And in fact...

- ◆ 18 March 2011...
- ◆ **.xxx** has been approved....
- ◆ Real effect?
- ◆ 200 milion dollars a year more in the bank account of  
**ICM Registry (...!!)**

