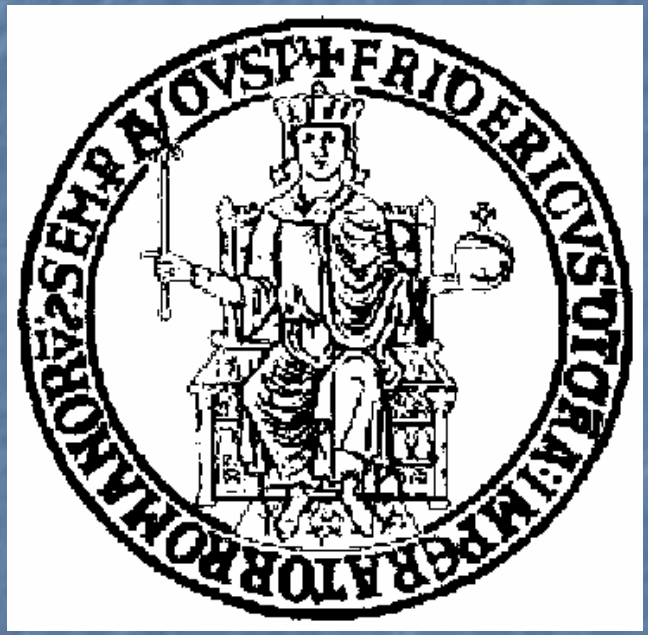
Improving Usability of Web Pages for Blind



Porfirio Tramontana
Carmine Cesarano
Anna Rita Fasolino

Dipartimento di Informatica e Sistemistica
University of Naples Federico II, Italy

Accessibility of Web Sites

- The World Wide Web has a growing importance in the life of anyone, in particular of people with phisical disabilities
  - Web accessibility is a fundamental issue for the development of today's Web sites and applications, stated by the most important international governments and organisations

Accessibility for Blind Users

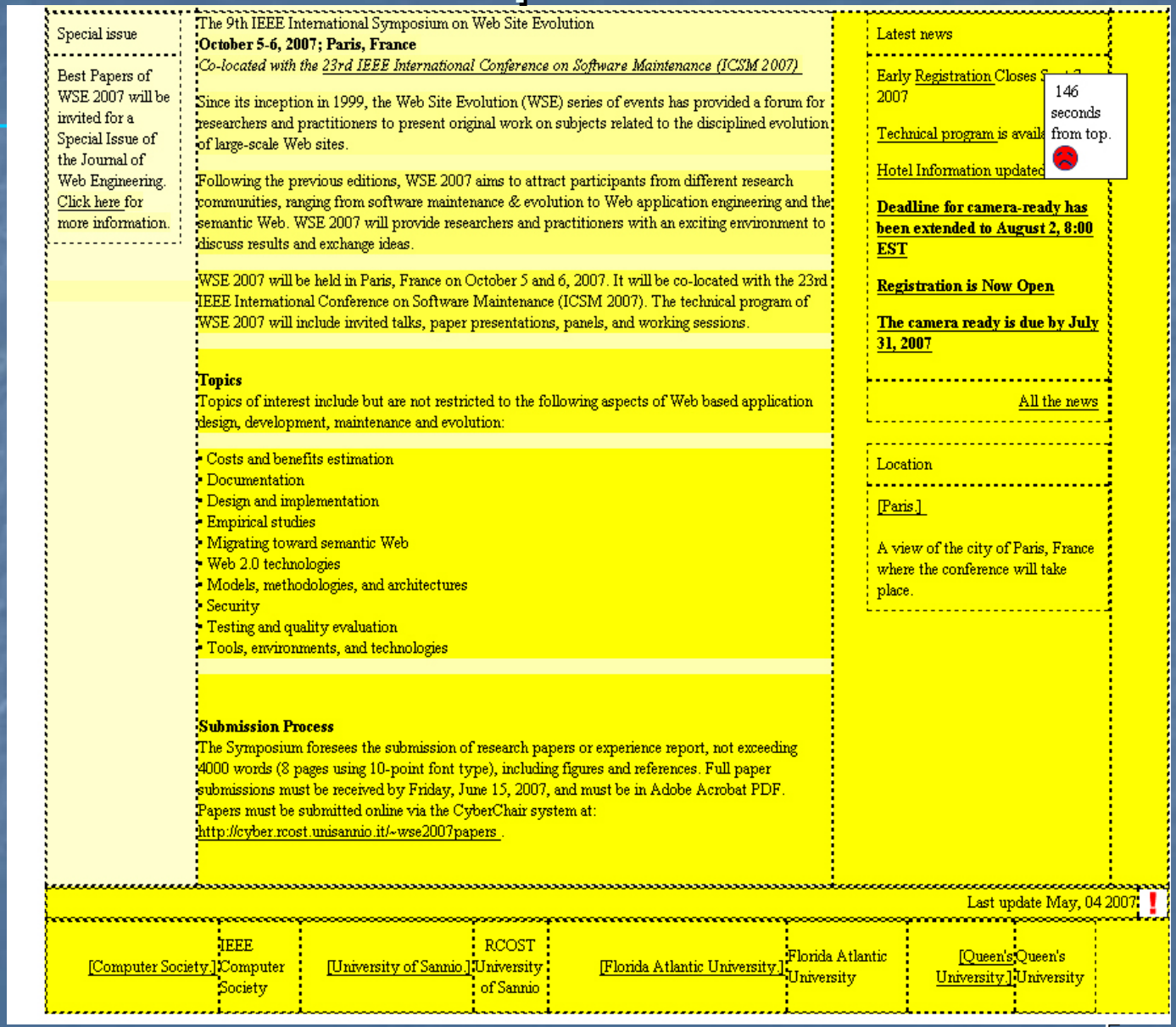
- Blind users can access textual contents of Web Sites by using Screen Readers
  - The blind user has to listen the full textual content of a Web page in order to obtain the needed information
  - Aural Web Sites are optimised for blind users
    - Bolchini et al. stated in 2006 some requirements to take into account for designing Aural Web Sites

Web Site Usability for Blinds

- The most part of the existing Web sites are not Aural Web Sites
  - Blind users experience accessibility and usability problems
    - The Web Sites are organised according to a bi-dimensional layout, while Screen Readers and Braille Readers provide the Web site information just in a linear way, as a mono-directional text
    - The time needed to access some contents can be very long, so that the Web site is not usable for blinds

An example

- IBM ADesigner visually reports the time needed to access information in a Web page.
- The time needed to access Submission information and recent news is more than two minutes!



Auralisation of Existing Web Sites

- In this work, we propose a solution to improve the usability of existing Web Sites, by generating ‘on the fly’ a summarised version of client pages
- Requirements of the solution:
  1. To minimise the reaching time i.e. the time required to the blind user to get the information which he/she is interested in;
  2. To minimise the computational time and resources needed to transform the original Web page in the auralised one;
  3. To be potentially applicable to any Web page, independently from its structure, layout and topics;
  4. To be tuneable, so that the usability of the auralised page could be further improved by tuning some parameters.

Summarisation Techniques

- Two different summarisation techniques have been proposed:

- T1) Based on structural analysis of built client pages;

- T2) Based on the analysis of the textual contents of built client pages

Technique T1 (based on Structural Analysis)

- Two heuristics have been taken into account for weighting the Web page information content importance:

  H1) the most important information are highlighted or emphasised by means of particular editing formats, in order to catalyse the attention of the user;

  H2) the most important information are usually placed in the higher parts of a Web page (to make them earlier visible to users);

Example

- H2) Contents in the higher part of the Built Client Page
- H1) Enhanced Contents

Weighting the contents: heuristic H1

- Any Text Segment (text fragment enclosed in tags) is weighted by taking into account the enhancing effect of the nested tags that influence the visualisation of the text

$$TW(TS) = \Pi_{t \in NS(T_i)} [W(t) * W(TStyle(t))]$$

Hiding Tags: W<1

Neutral Tags: W=1

Enhancing Tags: W>1

Tag Name	Weight
S, DEL, A, SMALL, STRIKE, …….	0,25
KBD, H6, CITE, CODE, … ……	0,5
COL, COLGROUP, COMMENT, DD, H5, DEN, DIR, EMBED, EM, THEAD, BUTTON, TFOOT, FIELDSET, FN, FONT, FRAME, FRAMESET, TEXTAREA, DIV, ADDRESS, TT, U, UL, VAR, WBR, XMP, SERVER, SHADOW, SIDEBAR, BODY, ACRONYM, BR, HTML, ……	1
B, OL, DL, STRONG, MENU, H3, BIG, Q PRE, TH, TR, I, CENTER, CAPTION, FORM, ……	1,5
TITLE, H2, ……	2
H1	3

  Hello World

TW('Hello World') = W('b')*W('I') = 1.5*1.5= 2.25.

Weighting the contents: heuristic H2

- Any Text Segment is weighted by taking into account the position of the text in the page

Example: Linear position weighting functions

$$PW(TS) = 1 - \frac{Pos(TS)}{Lenght(P)}$$

The total Weight of any text content is given by the product of the weights:

$$R(TS) = TW(TS) * PW(TS)$$

Given a fixed threshold, a summarised version of the Built client page is obtained by selecting only the contents with a weight that is greater than the threshold

Technique T2 (based on content summarisation)

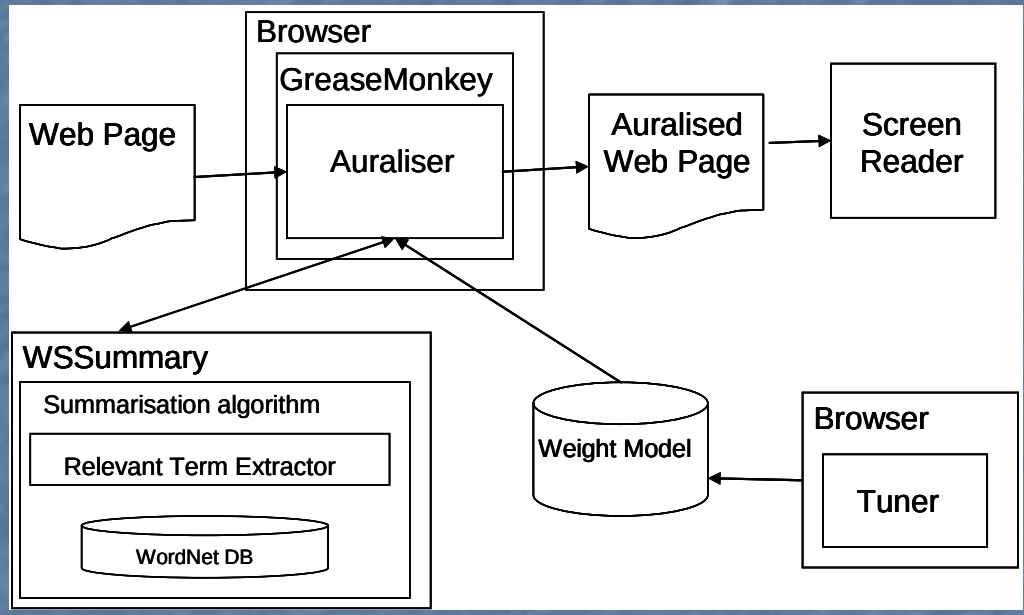
- A summary containing just sentences is obtained by taking into account the relative closeness (on the basis of a semantic net) of the terms they include. [*]

1. Nouns contained in the sentences are considered
2. A semantic net of this nouns is built by considering synonyms, hypernyms, hyponyms, meronyms
3. The relevance of a term in the page is evaluated by taking into account how much times the term appears in the page and how much times its synonims/hyponyms/meronyms appear
4. The relevance of a sentence is the sum of the relevance of its nouns
5. The summary is obtained by selecting the most relevant sentences

[*] P.Capasso, C. Cesarano, A. Picariello, L. Sansone, "Content-based News Retrieval on the Web", International Journal on Computer Science and Network Security, 2006, Vol.6, No.5A, pp.68-94

The Auralisation Framework

- The presented techniques have been implemented in a framework:
  - Weights needed for the technique T1 are maintained in a repository (Weight Model) and can be modified by the user via a Tuner client application



- The summarisation technique T2 is performed by a Web Service, with the support of WordNet
- The Auraliser is a Javascript client application, executed in the GreaseMonkey environment, generating the textual summary (Auralised Web Page) for the Screen Reader

Case Study

- Some explorative case studies have been carried out in order to assess the eventual increase of the usability of the Web pages

- Usability Measures
  - Let W be the Web page under experimentation and W* its auralised counterpart
  - Let RC be the set of pieces of information that the user expects to retrieve in the Web page and RC* $\subseteq$ RC the ones that are sufficiently reported in the auralised Web page

  - The *Relevant Concepts Recall* *RCR* = |RC*| / |RC|
  - The *Reaching Time* *RT* is the time after which a Screen Reader user can listen all the Relevant Concepts of RC* contained in W*.

Comparisons

- The proposed usability parameters have been evaluated with respect to:
  - the plain text contained in the Web page, without applying any transformation technique;
  - the auralised version obtained by applying the structural analysis technique (T1);
  - the auralised version obtained by applying the proposed summarisation technique (T2).

WSE Case Study

- 5 relevant concepts:
  - name and location of the conference (C1),
  - Special issue call (C2),
  - Introduction (C3),
  - Topics List (C4),
  - Submission Information (C5).

- Technique T1 was executed by fixing the minimum threshold for which the recall RCR=1

WSE 2007 Home Page

Results

Expected Concept	Adopted Technique		
	None	T1	T2
RT(C1)	29 sec	15 sec	-
RT(C2)	18 sec	5 sec	-
RT(C3)	68 sec	49 sec	21 sec
RT(C4)	118 sec	89 sec	-
RT(C5)	121 sec	93 sec	27 sec
RT (to reach all the page	121 sec	93 sec	-
Saved RT%	0	23	-
RCR	1	1	0.4

- In this example, Technique T1 reduces the Reaching Time value for any expected concept by applying
- Summarisation technique T2 includes just 2 concepts but they are accessed in a very short time

Discussion

- Although a larger experimentation is needed, the case studies that have been carried out assess that the two proposed summarisation techniques can be adopted to make a Web page more usable to blind users

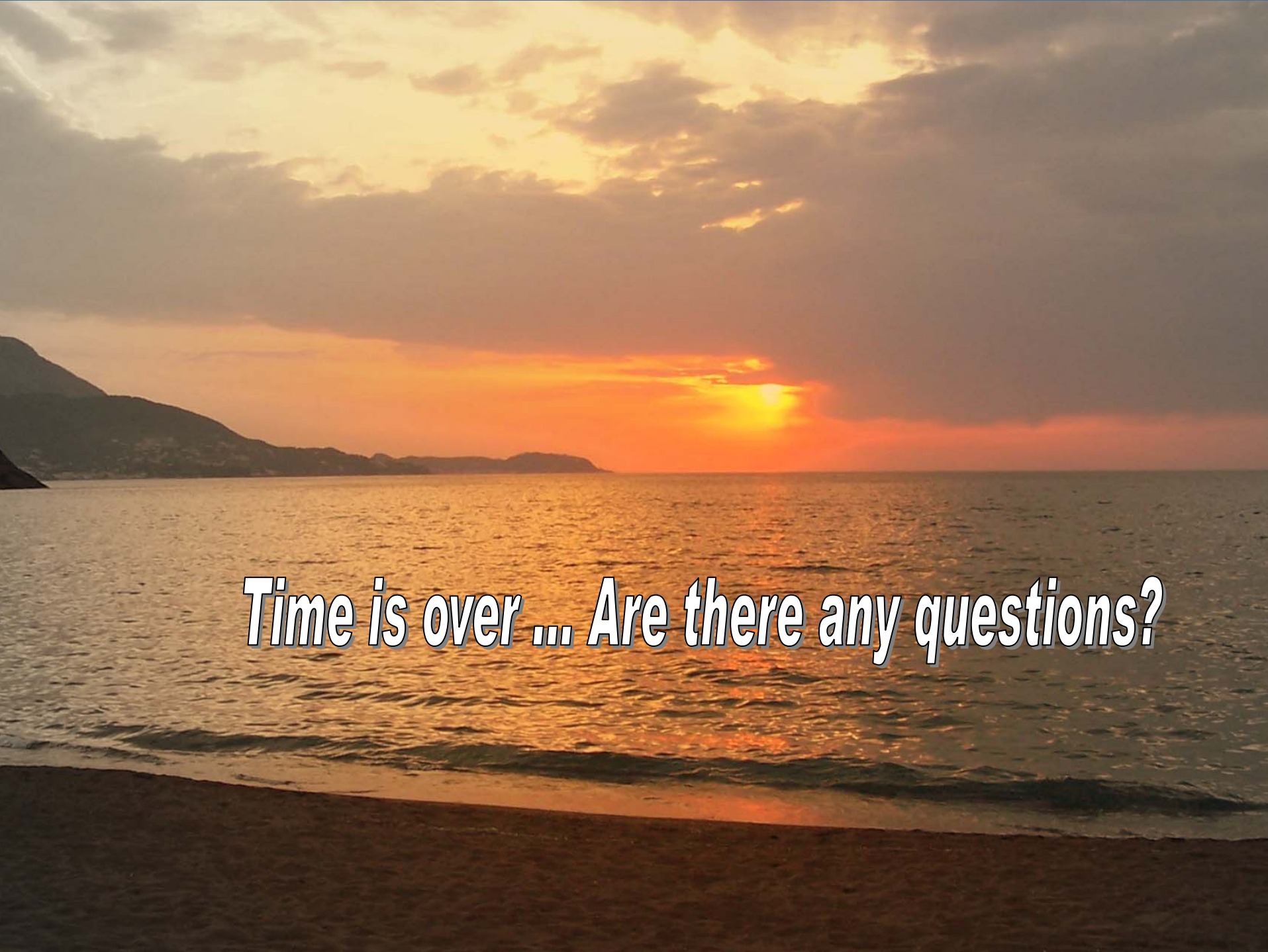
- Anyway, the two proposed techniques presents some limitations:
  - Technique T1 gives good results if weight models comprehending specific style weights, too, are adopted
    - Weight models could be provided by Web site publisher
  - Technique T2 could be, sometimes too computational expensive, with respect to expected reaching time

Conclusions

- Two different summarisation techniques have been proposed for the generation of Auralised Web pages that reduces the Reaching Time needed to blind users to access contents of a Web Site

Future Works

- Some ideas:
  - Reaching times can be furtherly reduced if the Auraliser inserts internal anchors to the more relevant contents in the built summarised pages
  - Considering hybrid techniques combining T1 and T2
  - Generalising the proposed approach to client pages which contents are dynamically generated (RIAs)


Time is over ... Are there any questions?