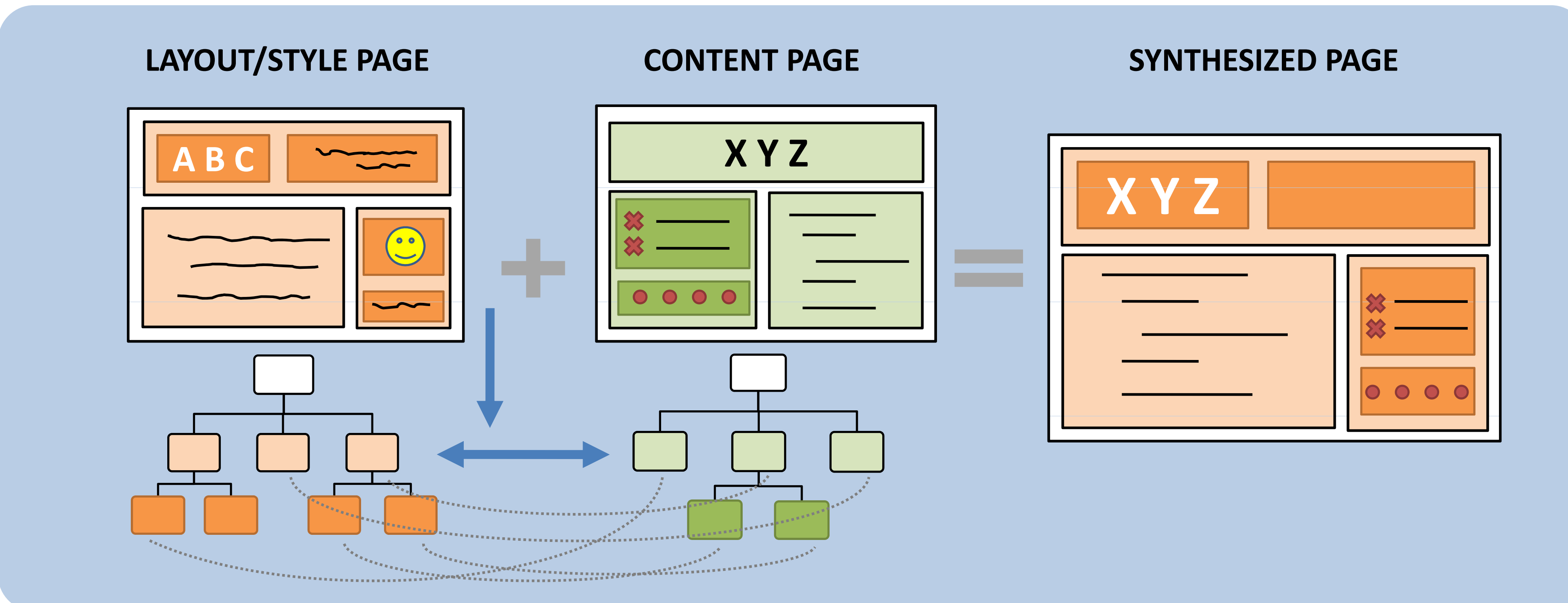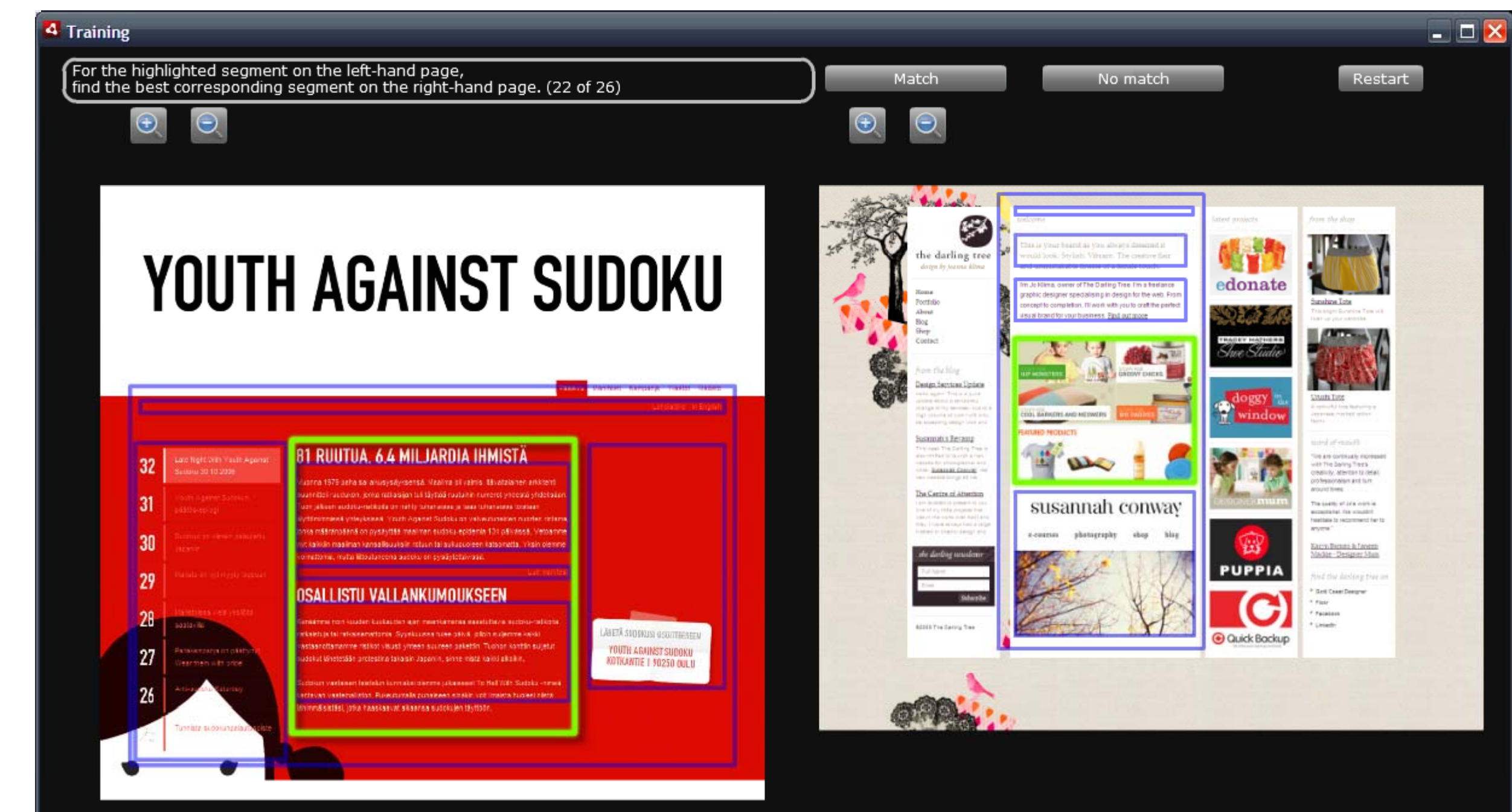Juho Kim, Ranjitha Kumar, Scott R. Klemmer

Crowdsourcing Interface for Collecting Correspondences of Web Pages

The more than **one-trillion** pages on the web today provide a corpus of design examples **unparalleled in human history**. How can we leverage these examples to create new designs?
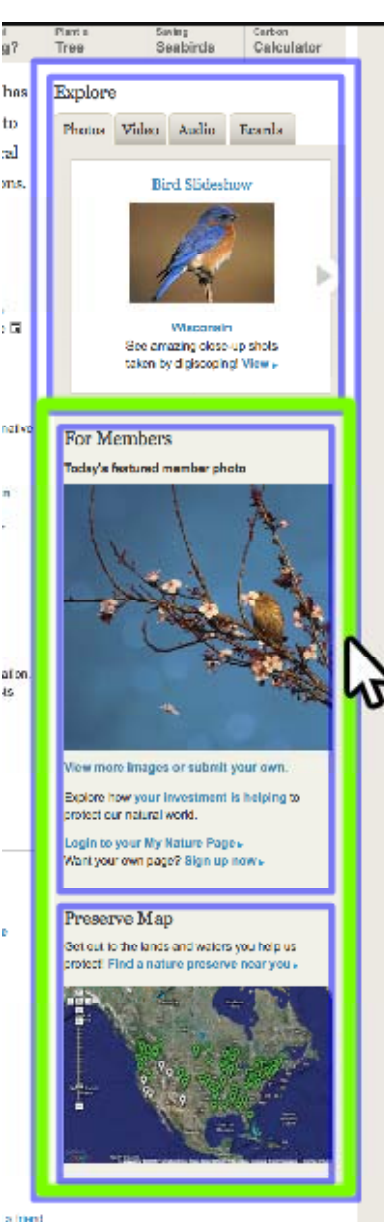

LAYOUT/STYLE PAGE   CONTENT PAGE   SYNTHESIZED PAGE

Computer Vision   decompose web pages into visual blocks

Machine Learning   compute semantic correspondences between pages

Crowdsourcing   train on thousands of mappings from hundreds of users

ORIGINAL PAGE   ALTERNATIVE DESIGN #1   ALTERNATIVE DESIGN #2

Design Decisions

Serialization

Minimize User Errors
• Minimize accidents: Two-click interface
• Review previous matches: Brushing and linking
• Precise selection: Bloating parent nodes

Provide Relevant Information
• Gain familiarity: 30-second page preview
• Reduce ordering bias: randomized depth-first ordering
• Hints: locally-related nodes shown on mouse-over

Results

Preliminary Evaluation

Experiment	Authors	MTurk 1	MTurk 2
Mappings assigned per user	210	5	1
No. of users	2	40	40
Total mappings collected	420	200	40
Task duration (sec)	156.8	277.1	340.6
Nodes mapped (%)	91.5	71.2	71.4

Turkers take longer and define fewer mappings

Future Evaluation
• Generate synthesized pages based on mappings
• Experts rate mapping quality based on the synthesized result