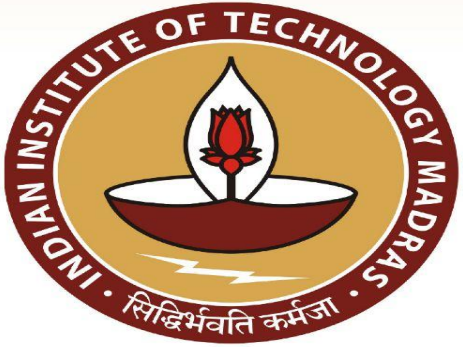


IIT Madras
BSc Degree

Markup

Markup

- Information representation
- Raw data vs Semantics
- Logical structure vs Styling
- HTML5 and CSS

Information representation

- Computers work only with “bits”
  - Binary digits: 0 and 1
- Numbers
  - Place value: binary numbers: eg. 6 = 0110
  - Two’s complement: negative numbers: eg. -6 = 1010
- Letters? Arbitrary Text?

Representing Text

- ASCII
- Unicode
- UTF-8

Information Interchange

- Communicate through machines - either between machines or between humans
- Machines only work with ***bits***
- Standard “encoding”
    - Some sequence of bits interpreted as a character

Interpretation

What is “0100 0001”?

- String of bits
- Number with value 65 decimal
- Character “A”
- All of the above

Interpretation

What is “0100 0001”?

- String of bits
- Number with value 65 decimal
- Character “A”
- All of the above

Matter of **interpretation** and **Context**

ASCII

- American Standard Code for Information Interchange
- 7-bits: 128 different entities
  - ‘a’ .. ‘z’ - 26
  - ‘A’ .. ‘Z’ - 26
  - ‘0’ .. ‘9’ - 10
  - Special characters: !@#$%^&*() ...

  } 126
- Why 7-bits?
- What about other characters? अ அ ಅ ജ 不
  - 1000s of characters needed

Unicode

- Allow codes for more scripts, characters
- How many?
  - All living languages? All extinct languages? All future languages?
- “Universal Character Set” encoding - UCS
  - UCS-2: 2 bytes per character - max 65,536 characters → 2^16
  - UCS-4: 4 bytes per character: 4 Billion+ characters → 2^32

Efficiency?

- Most common language on Web: ???
- Should all characters be represented with same number of bits?

Efficiency?

- Most common language on Web: ???
- Should all characters be represented with same number of bits?
- Example:
  - Text document with 1000 words, approximately 5000 characters (including spaces)

Efficiency?

- Most common language on Web: ???
- Should all characters be represented with same number of bits?
- Example:
  - Text document with 1000 words, approximately 5000 characters (including spaces)
  - UCS-4 encoding: 32b x 5000 = 160,000 bits

Efficiency?

- Most common language on Web: ???
- Should all characters be represented with same number of bits?
- Example:
  - Text document with 1000 words, approximately 5000 characters (including spaces)
  - UCS-4 encoding: 32b x 5000 = 160,000 bits
  - ASCII encoding: 8b x 5000 = 40,000 bits
  - Original 7-bit ASCII sufficient for English: 7b x 5000 = 35,000 bits

Efficiency?

- Most common language on Web: ???
- Should all characters be represented with same number of bits?
- Example:
  - Text document with 1000 words, approximately 5000 characters (including spaces)
  - UCS-4 encoding: 32b x 5000 = 160,000 bits
  - ASCII encoding: 8b x 5000 = 40,000 bits
  - Original 7-bit ASCII sufficient for English: 7b x 5000 = 35,000 bits
  - Minimum needed to encode just ‘a’ - ‘z’, numbers and some special characters: could fit in 6 bits: 30,000 bits

Efficiency?

- Most common language on Web: ???
- Should all characters be represented with same number of bits?
- Example:
  - Text document with 1000 words, approximately 5000 characters (including spaces)
  - UCS-4 encoding: 32b x 5000 = 160,000 bits
  - ASCII encoding: 8b x 5000 = 40,000 bits
  - Original 7-bit ASCII sufficient for English: 7b x 5000 = 35,000 bits
  - Minimum needed to encode just ‘a’ - ‘z’, numbers and some special characters: could fit in 6 bits: 30,000 bits
  - Optimal coding based on frequency of occurrence:
    - ‘e’ is most common letter, ‘t’, ‘a’, ‘o’, ...
    - Huffman or similar encoding: ~ 10-20,000 bits, possibly less

Solvable in general?

- Impossible to encode by actual character frequency: depends on text
  - Just use compression methods like “zip” instead!
- But can encoding be a good halfway point?

Example:

- Use 1 byte for most common alphabets
- Group others according to frequency, have “prefix” codes to indicate

Prefix Coding

1st Byte	2nd Byte	3rd Byte	4th Byte	Free Bits	Maximum Expressible Unicode Value
0xxxxxxx				7	007F hex (127)
110xxxxx	10xxxxxx			(5+6)=11	07FF hex (2047)
1110xxxx	10xxxxxx	10xxxxxx		(4+6+6)=16	FFFF hex (65535)
11110xxx	10xxxxxx	10xxxxxx	10xxxxxx	(3+6+6+6)=21	10FFFF hex (1,114,111)

Example

	A	א	好	不
Code point	U+0041	U+05D0	U+597D	U+233B4
UTF-8	41	D7 90	E5 A5 BD	F0 A3 8E B4
UTF-16	00 41	05 D0	59 7D	D8 4C DF B4
UTF-32	00 00 00 41	00 00 05 D0	00 00 59 7D	00 02 33 B4

Src: https://www.w3.org/International/articles/definitions-characters/

UTF-8

- Use 8 bits for most common characters: ASCII subset
  - All ASCII documents are automatically UTF-8 compatible
- All other characters can be encoded based on prefix string
- More difficult for text processor:
  - first check prefix
  - linked list through chain of prefixes possible
  - Still more efficient for majority of documents
- Most common encoding in use today

Markup

- Content vs Meaning
- Types of markup
- (X)HTML

Content

Markup What is markup? Markup is a way of using cues or codes in the regular flow of text to indicate how text should be displayed. Markup is very useful to make the display of text clear and easy to understand.

Markup

Title

Heading level 1

Markup What is markup? Markup is a way of using cues or codes in the regular flow of text to indicate how text should be displayed. Markup is very useful to make the display of text clear and easy to understand.

Insert para break

Result

Markup

What is markup?

Markup is a way of using cues or codes in the regular flow of text to indicate how text should be displayed.

Markup is very useful to make the display of text clear and easy to understand.

Types of Markup

Coombs et al, “Communication Systems and the Future of Scholarly Text Processing”, Communications of ACM, 1987

Types of Markup

- Presentational
  - WYSIWYG: directly format output and display
  - Embed codes not part of regular text, specific to the editor

Coombs et al, “Communication Systems and the Future of Scholarly Text Processing”, Communications of ACM, 1987

Types of Markup

- Presentational
  - WYSIWYG: directly format output and display
  - Embed codes not part of regular text, specific to the editor
- Procedural
  - Details on how to display:
    - change font to large, bold
    - skip 2 lines, indent 4 columns

Coombs et al, “Communication Systems and the Future of Scholarly Text Processing”, Communications of ACM, 1987

Types of Markup

- Presentational
  - WYSIWYG: directly format output and display
  - Embed codes not part of regular text, specific to the editor
- Procedural
  - Details on how to display:
    - change font to large, bold
    - skip 2 lines, indent 4 columns
- Descriptive
  - This is a <title>, this is a <heading>, this is a <paragraph>

Coombs et al, “Communication Systems and the Future of Scholarly Text Processing”, Communications of ACM, 1987

Examples

- MS Word, Google Docs etc:
  - User interface focused on “appearance”, not meaning
  - WYSIWYG: direct control over styling
  - Often leads to complex formatting and loss of inherent meaning
- LaTeX, HTML (general *ML)
  - Focus on meaning
  - More complex to write and edit, not WYSIWYG in general

Semantic Markup

- **Content** vs **Presentation**
- *Semantics*
  - Meaning of the text
  - structure or logic of the document

HTML (and co.)

- HyperText Markup Language
- Generalizations
- Variants of Interest

HyperText Markup Language

- HTML first used by Tim Berners-Lee in original Web at CERN (~1989)
- Considered an *application* of **SGML** (Standard Generalized Markup Language)
  - Strict definitions on structure, syntax, validity
- HTML meant for browser interpretation
  - Very forgiving: loose validity checks
  - Best effort to display

HTML Example

```
<!DOCTYPE html>
<html>
<body>

<h1>My First Heading</h1>

<p>My first paragraph.</p>

</body>
</html>
```

Tags

- <h1> </h1> - paired tags
- Angle brackets < >
- Closing tag with /
- Location specific: <DOCTYPE>: only at head of doc
- Case-insensitive

Nesting

- Hello
- ***Hello***

Invalid:

- Hello
- Hello
- Hell<o/em>

Presentation vs Semantics

- Hello
- Hello
- **Hello**

Which one is right? Which is better?

Timelines

- SGML based
  - 1989 - HTML original
  - 1995 - HTML 2
  - 1997 - HTML 3, 4
- XML based
  - XHTML - 1997 - mid 2010s
- HTML5
  - first release 2008
  - W3C recommendation - 2014

HTML5

- Block elements:
- Inline elements:
- Logical elements: <nav>, <footer>
- Media: <audio>, <video>

Remove “presentation only” tags:

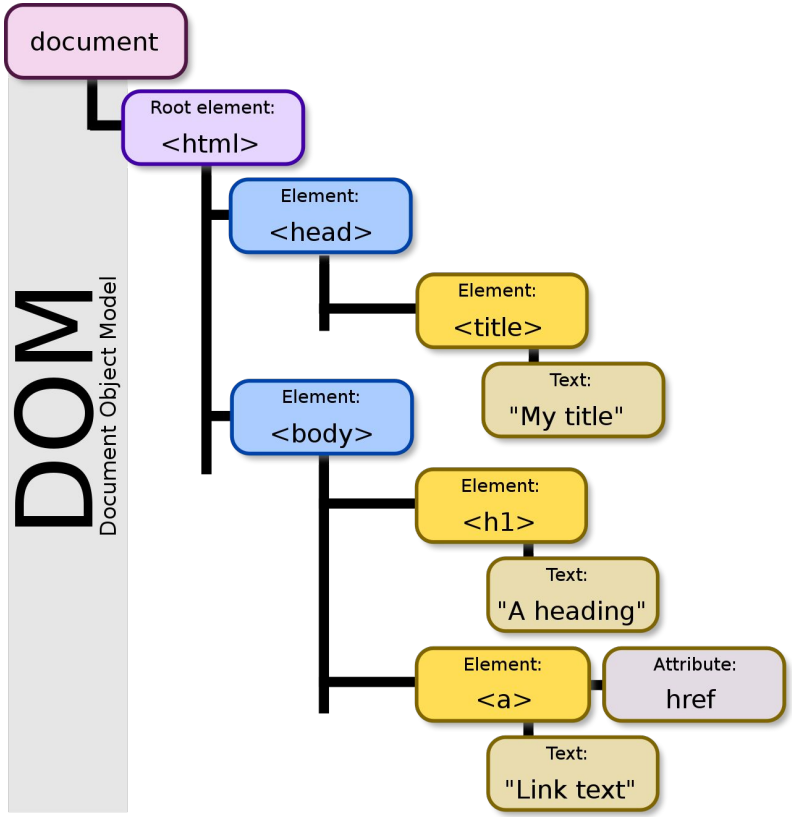
-
- <font>

Document Object Model

```
<html>
<head>
    <title>My title</title>
</head>
<body>
    <h1>A heading</h1>
    <a href=”link”>Link Text</a>
</body>
</html>
```

Document Object Model

```
<html>
<head>
    <title>My title</title>
</head>
<body>
    <h1>A heading</h1>
    <a href=”link”>Link Text</a>
</body>
</html>
```



Src: B. Eriksson, Wikipedia

DOM

- Tree structure representing logical layout of document
- *Direct manipulation of tree possible!*
- Application Programming Interfaces (APIs)
  - Canvas
  - Offline
  - Web Storage
  - Drag and Drop
  - ...
- Javascript primary means of manipulating
- CSS used for styling

Styling

- Markup vs Style
- Themes
- CSS

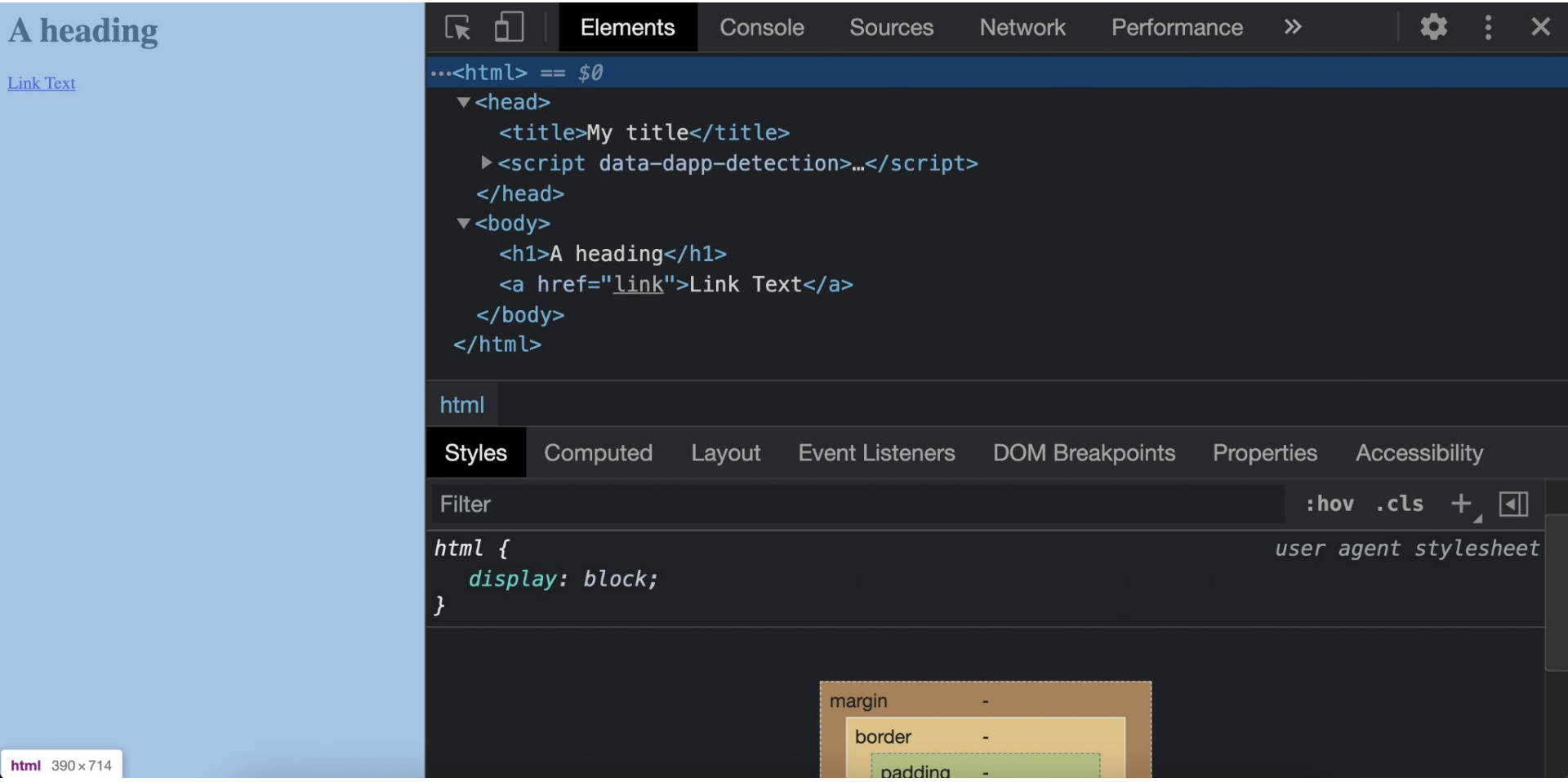
Markup vs Style

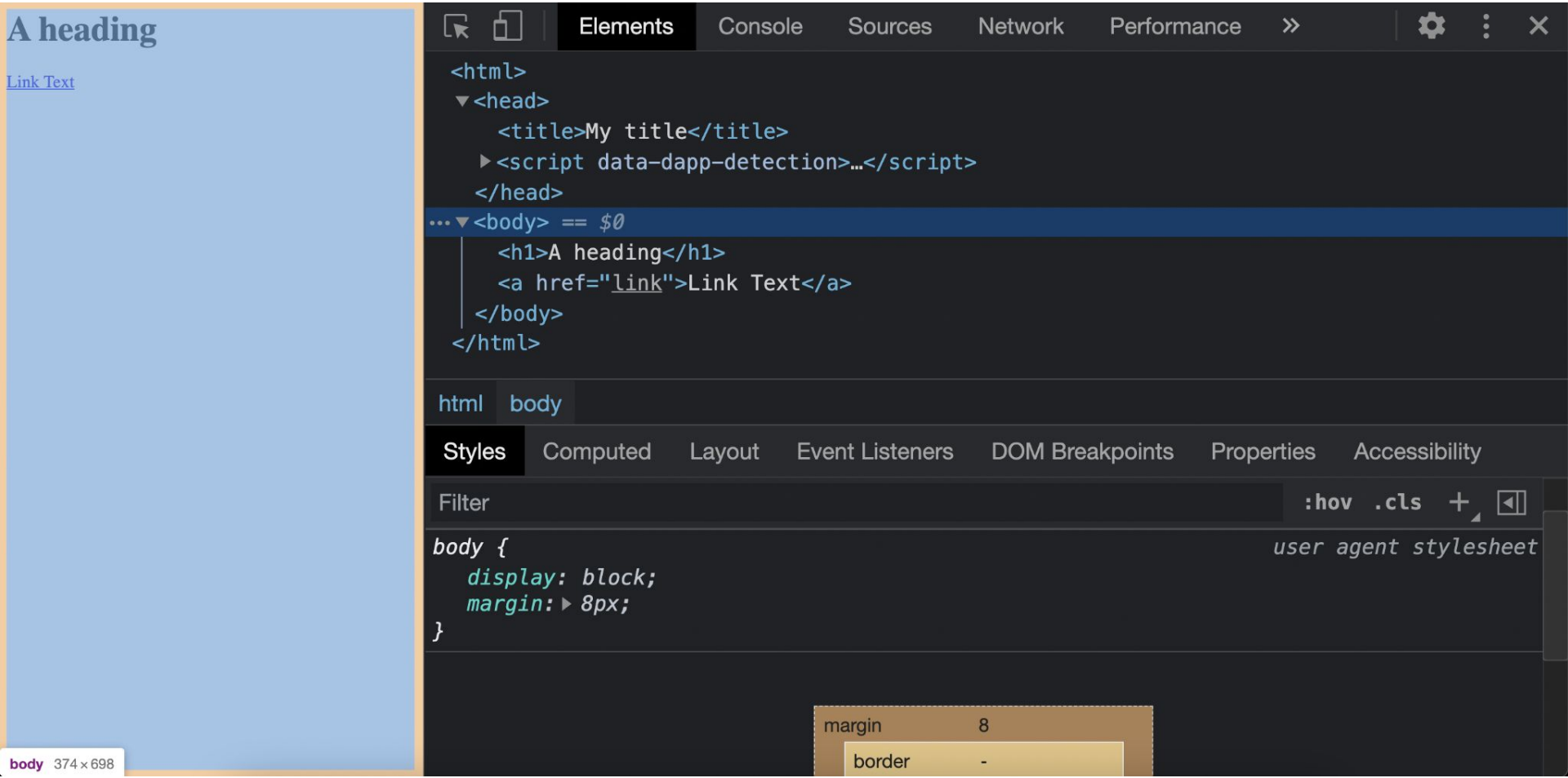
<h1>Hello</h1>

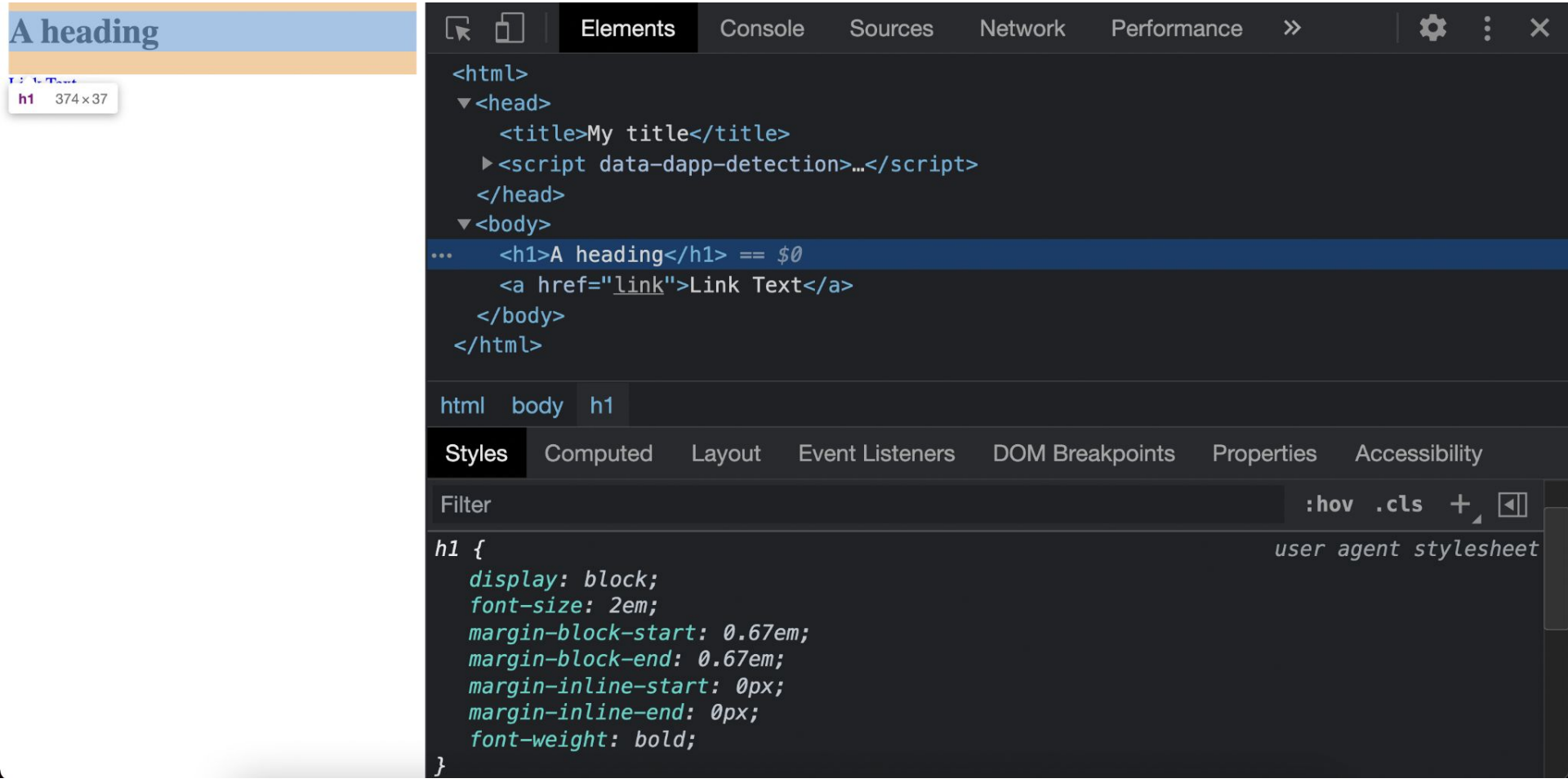
Hello	Font - Garamond, Size: 24, Bold
Hello	Font: Arial, Size: 30, Bold
Hello	Font: Comic Sans, Size: 24, Bold, Italic, FontColor: Green, Background: Red

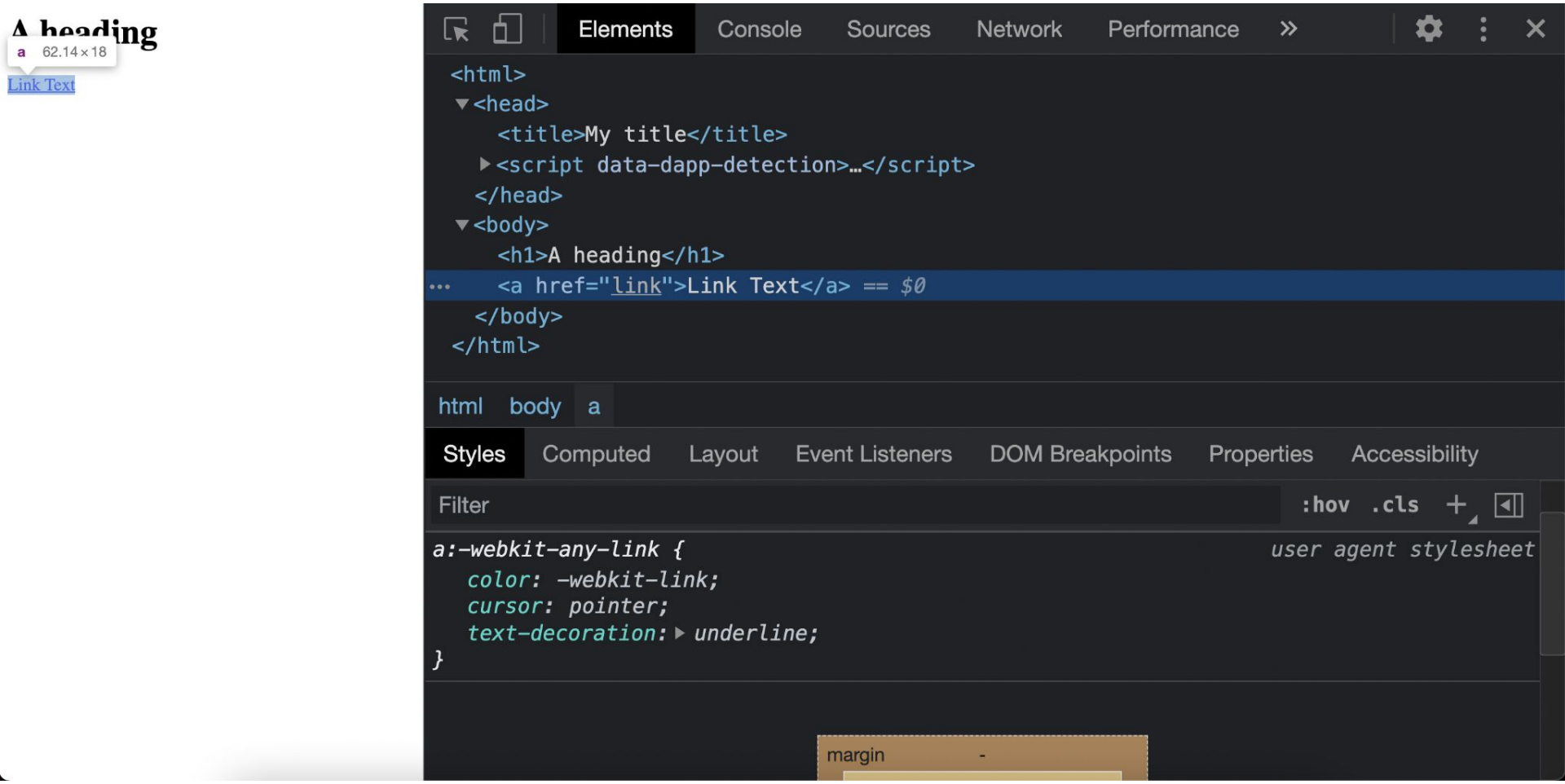
Separation of Styling

- Style hints in separate blocks
  - Separate files included
- Themes
- Style Sheets
  - Specify presentation information
- Cascading Style Sheets (CSS)
  - Allow multiple definitions
  - Latest takes precedence


A heading
Link Text
html 390 × 714
Elements
Console
Sources
Network
Performance
<html> == $0
<head>
<title>My title</title>
<script data-dapp-detection>…</script>
</head>
<body>
<h1>A heading</h1>
Link Text
</body>
</html>
html
Styles
Computed
Layout
Event Listeners
DOM Breakpoints
Properties
Accessibility
Filter
:hov .cls
html {
display: block;
}
user agent stylesheet
margin -
border -
padding -


A heading
Link Text
body 374 × 698
Elements
Console
Sources
Network
Performance
<html>
<head>
<title>My title</title>
<script data-dapp-detection>…</script>
</head>
<body> == $0
<h1>A heading</h1>
Link Text
</body>
</html>
html
body
Styles
Computed
Layout
Event Listeners
DOM Breakpoints
Properties
Accessibility
Filter
:hov .cls
body {
display: block;
margin: 8px;
}
user agent stylesheet
margin 8
border -

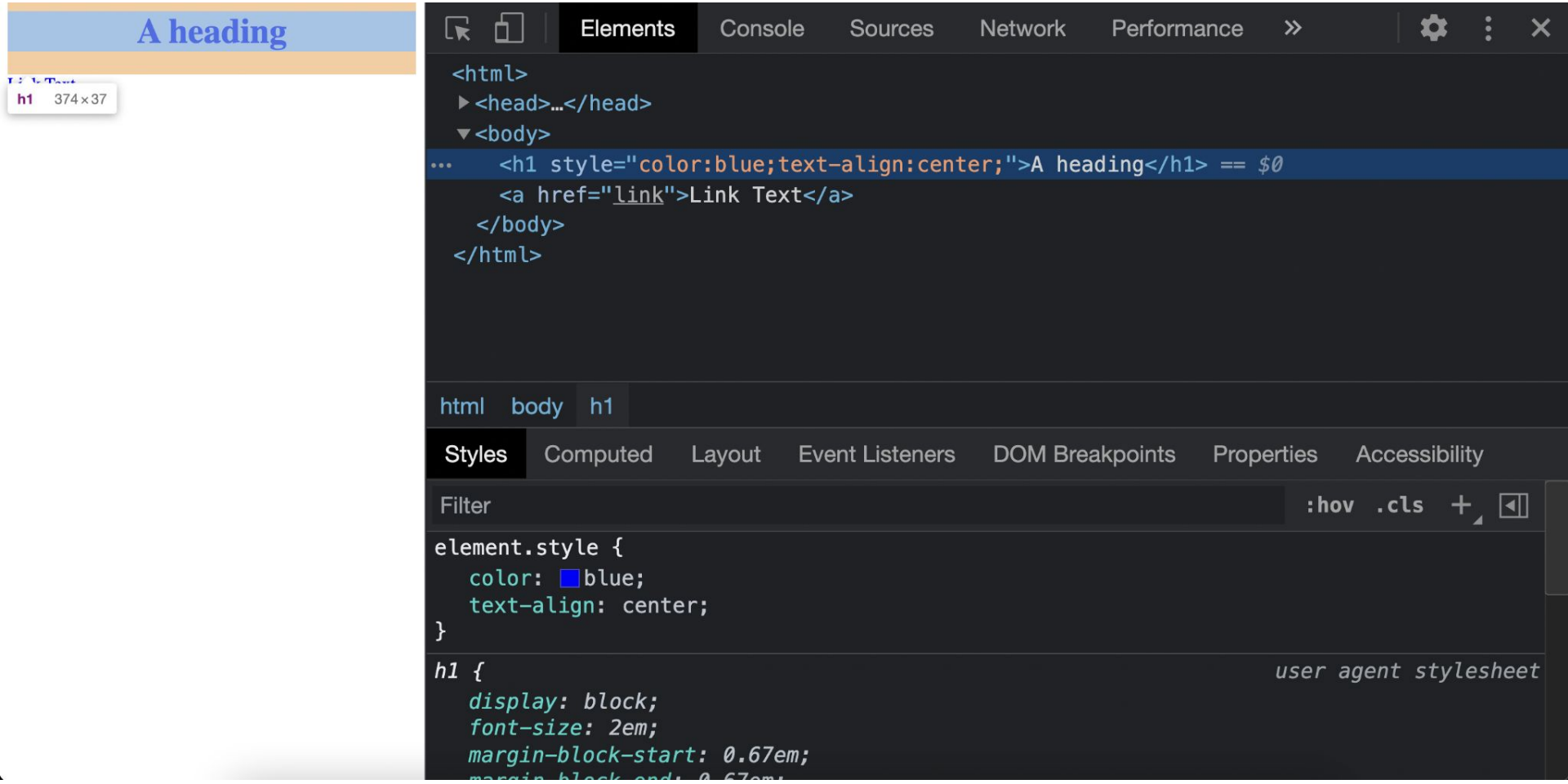

A heading
h1 374 × 37
Elements
Console
Sources
Network
Performance
<html>
<head>
<title>My title</title>
<script data-dapp-detection>…</script>
</head>
<body>
<h1>A heading</h1> == $0
Link Text
</body>
</html>
html body h1
Styles
Computed
Layout
Event Listeners
DOM Breakpoints
Properties
Accessibility
Filter
:hov .cls
h1 {
user agent stylesheet
display: block;
font-size: 2em;
margin-block-start: 0.67em;
margin-block-end: 0.67em;
margin-inline-start: 0px;
margin-inline-end: 0px;
font-weight: bold;
}


A heading
a 62.14×18
Link Text
Elements
Console
Sources
Network
Performance
<html>
<head>
<title>My title</title>
<script data-dapp-detection>…</script>
</head>
<body>
<h1>A heading</h1>
Link Text == $0
</body>
</html>
html body a
Styles
Computed
Layout
Event Listeners
DOM Breakpoints
Properties
Accessibility
Filter
:hov .cls
a:-webkit-any-link {
user agent stylesheet
color: -webkit-link;
cursor: pointer;
text-decoration: underline;
}
margin -

Inline CSS

- Directly add style to the tag
- Example:

```
<h1 style="color:blue;text-align:center;">A heading</h1>
```

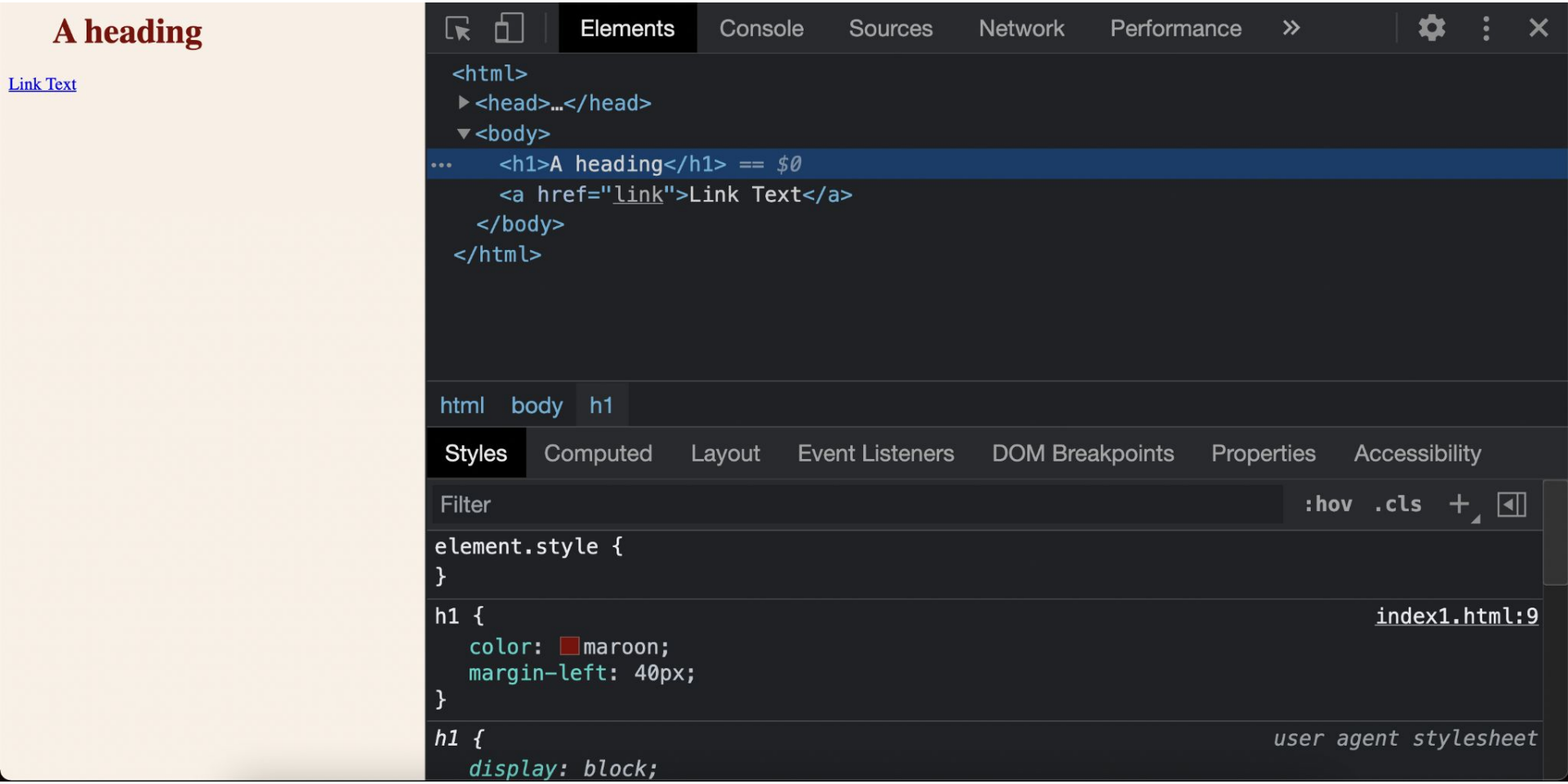

A heading
h1 374 × 37
Elements
Console
Sources
Network
Performance
<html>
<head>…</head>
<body>
<h1 style="color:blue;text-align:center;">A heading</h1> == $0
Link Text
</body>
</html>
html
body
h1
Styles
Computed
Layout
Event Listeners
DOM Breakpoints
Properties
Accessibility
Filter
:hov
.cls
element.style {
color: blue;
text-align: center;
}
h1 {
user agent stylesheet
display: block;
font-size: 2em;
margin-block-start: 0.67em;

Internal CSS

- Embed inside <head> tag
- Now all <h1> tags in document will look the same - centrally modified

```
<style>
body {
  background-color: linen;
}

h1 {
  color: maroon;
  margin-left: 40px;
}
</style>
```


A heading
Link Text
Elements
Console
Sources
Network
Performance
<html>
<head>…</head>
<body>
<h1>A heading</h1> == $0
Link Text
</body>
</html>
html
body
h1
Styles
Computed
Layout
Event Listeners
DOM Breakpoints
Properties
Accessibility
Filter
:hov
.cls
element.style {
}
h1 {
index1.html:9
color: maroon;
margin-left: 40px;
}
h1 {
user agent stylesheet
display: block;

External CSS

- Extract common content for reuse
- Multiple CSS files can be included
- Latest definition of style takes precedence

Responsive Design

- Mobile and Tablets have smaller screens
  - Different form factors
- Adapt to screen - ***Respond***
- CSS control styling - HTML controls content!

Department Of Electrical Engineering

Indian Institute of Technology Madras

News

MV Scholar 2020

EE Department / November 18, 2020

Ms. Lakshmi Areekath (EE17D014) is the winner of the Malathi Veeraraghavan Scholar 2020 Award instituted in the memory of...

Faculty recognition

EE Department / November 18, 2020

Dr. Sheetal Kalyani, has been appointed on her following Editorial Board Memberships : Associate Editor, IEEE Trans on Signal Processing (since October 2020)...

MS/PhD Admission 2021-22

List of MS & Ph.D. shortlisted candidates for interview

eewebmaster / May 17, 2021

The interviews will be held online. Interview dates are as follows: EE1, EE2, EE4, EE5, and EE6 : 31...

M.S. / Ph.D. Admission Criteria July 2021

eewebmaster / May 16, 2021

UG% means the percentage marks scored in your 4 year UG programme, as reported by you, rounded up to...

Upcoming Events Featured

Third National Power Engineering Research Scholars' Conference (NPERSC) at IIT Madras, during 11-12 September 2021

September 11 @ 8:00 am - September 12 @ 5:00 pm

View All Events

Search

Search

Opportunities

MV Scholar Program 2020 @ IITM

HappEE

Celebrating student achievements:

Department Of Electrical Engineering

Indian Institute of Technology Madras

News

MV Scholar 2020

EE Department / November 18, 2020

Ms. Lakshmi Areekath (EE17D014) is the winner of the Malathi Veeraraghavan Scholar 2020 Award instituted in the memory of...

Faculty recognition

EE Department / November 18, 2020

Dr. Sheetal Kalyani, has been appointed on her following Editorial Board Memberships : Associate Editor, IEEE Trans on Signal Processing (since October 2020)...

MS/PhD Admission 2021-22

List of MS & Ph.D. shortlisted candidates for interview

eewebmaster / May 17, 2021

The interviews will be held online. Interview dates are as follows: EE1, EE2, EE4, EE5, and EE6 : 31...

M.S. / Ph.D. Admission Criteria July 2021

eewebmaster / May 16, 2021

UG% means the percentage marks scored in your 4 year UG programme, as reported by you, rounded up to...

Upcoming Events | Featured

Third National Power Engineering Research Scholars' Conference (NPERSC) at IIT Madras, during 11-12 September 2021
September 11 @ 8:00 am - September 12 @ 5:00 pm

View All Events

Search

Search

Opportunities

MV Scholar Program 2020 @ IITM

eewebmaster / August 7, 2020

MV Scholar Program 2020 @ IITM General Information The Malathi Veeraraghavan (MV) Scholar Program is a fellowship in the...

Faculty Positions

eewebmaster / March 10, 2015

Overview We routinely accept and review applications for faculty positions from individuals with exceptional research record in any area of...

HappEE

Celebrating student achievements: Sruti S

Uday Khankhoje / November 4, 2020

Sruti. S is a PhD student in the Department, and has been awarded the prestigious Prime Minister's Research Fellowship...

Home | People | Research | Academics | Events | Reach Us | Internal

Department Of Electrical Engineering

Indian Institute of Technology Madras

News | Opportunities | Featured | HappEE | MS/PhD Admission 2021-22

You are here: Home

News

MV Scholar 2020

EE Department / November 18, 2020

Ms. Lakshmi Areekath (EE17D014) is the winner of the Malathi Veeraraghavan Scholar 2020 Award instituted in the memory of...

Faculty recognition

EE Department / November 18, 2020

Dr. Sheetal Kalyani, has been appointed on her following Editorial Board Memberships : Associate Editor, IEEE Trans on Signal Processing (since October 2020)...

MS/PhD Admission 2021-22

List of MS & Ph.D. shortlisted candidates for interview

eewebmaster / May 17, 2021

The interviews will be held online. Interview dates are as follows: EE1, EE2, EE4, EE5, and EE6 : 31...

M.S. / Ph.D. Admission Criteria July 2021

eewebmaster / May 16, 2021

UG% means the percentage marks scored in your 4 year UG programme, as reported by you, rounded up to...

Upcoming Events | Featured

Third National Power Engineering Research Scholars' Conference (NPERSC) at IIT Madras, during 11-12 September 2021
September 11 @ 8:00 am - September 12 @ 5:00 pm

View All Events

Search

Search

Opportunities

MV Scholar Program 2020 @ IITM

eewebmaster / August 7, 2020

MV Scholar Program 2020 @ IITM General Information The Malathi Veeraraghavan (MV) Scholar Program is a fellowship in the...

Faculty Positions

eewebmaster / March 10, 2015

Overview We routinely accept and review applications for faculty positions from individuals with exceptional research record in any area of...

HappEE

Celebrating student achievements: Sruti S

Uday Khankhoje / November 4, 2020

Sruti. S is a PhD student in the Department, and has been awarded the prestigious Prime Minister's Research Fellowship...

Select Page: Where to?

Select Category: Where to?

Department Of Electrical Engineering

Indian Institute of Technology Madras

You are here: Home

News

MV Scholar 2020

EE Department / November 18, 2020

Ms. Lakshmi Areekath (EE17D014) is the winner of the Malathi Veeraraghavan Scholar 2020 Award instituted in the memory of...

Faculty recognition

EE Department / November 18, 2020

Dr. Sheetal Kalyani, has been appointed on her following Editorial Board Memberships : Associate Editor, IEEE Trans on Signal Processing (since October 2020)...

MS/PhD Admission 2021-22

List of MS & Ph.D. shortlisted candidates for interview

eewebmaster / May 17, 2021

The interviews will be held online. Interview dates are as follows: EE1, EE2, EE4, EE5, and EE6 : 31...

M.S. / Ph.D. Admission Criteria July 2021

eewebmaster / May 16, 2021

UG% means the percentage marks scored in your 4 year UG programme, as reported by you, rounded up to...

Opportunities

MV Scholar Program 2020 @ IITM

Select Page: Where to?

Select Category: Where to?

Department Of Electrical Engineering

Indian Institute of Technology Madras

You are here: Home

News

MV Scholar 2020

EE Department / November 18, 2020

Ms. Lakshmi Areekath (EE17D014) is the winner of the Malathi Veeraraghavan Scholar 2020 Award instituted in the memory of...

Faculty recognition

EE Department / November 18, 2020

Dr. Sheetal Kalyani, has been appointed on her following Editorial Board Memberships : Associate Editor, IEEE Trans on Signal Processing (since October 2020)...

MS/PhD Admission 2021-22

List of MS & Ph.D. shortlisted candidates for interview

eewebmaster / May 17, 2021

The interviews will be held online. Interview dates are as follows: EE1, EE2, EE4, EE5, and EE6 : 31...

M.S. / Ph.D. Admission Criteria July 2021

eewebmaster / May 16, 2021

UG% means the percentage marks scored in your 4 year UG programme, as reported by you, rounded up to...

Opportunities

MV Scholar Program 2020 @ IITM

Select Page: Where to?

Select Category: Where to?

Department Of Electrical Engineering

Indian Institute of Technology Madras

You are here: Home

News

MV Scholar 2020

EE Department / November 18, 2020

Ms. Lakshmi Areekath (EE17D014) is the winner of the Malathi Veeraraghavan Scholar 2020 Award instituted in the memory of...

Faculty recognition

EE Department / November 18, 2020

Dr. Sheetal Kalyani, has been appointed on her following Editorial Board Memberships : Associate Editor, IEEE Trans on Signal Processing (since October 2020)...

MS/PhD Admission 2021-22

List of MS & Ph.D. shortlisted candidates for interview

eewebmaster / May 17, 2021

The interviews will be held online. Interview dates are as follows: EE1, EE2, EE4, EE5, and EE6 : 31...

M.S. / Ph.D. Admission Criteria July 2021

eewebmaster / May 16, 2021

UG% means the percentage marks scored in your 4 year UG programme, as reported by you, rounded up to...

Opportunities

MV Scholar Program 2020 @ IITM

Select Page: Where to?

Select Category: Where to?

Department Of Electrical Engineering

Indian Institute of Technology Madras

You are here: Home

News

MV Scholar 2020

EE Department / November 18, 2020

Ms. Lakshmi Areekath (EE17D014) is the winner of the Malathi Veeraraghavan Scholar 2020 Award instituted in the memory of...

Faculty recognition

Bootstrap

- Commonly used framework
  - Originated from Twitter
  - Widely used now
- Standard styles for various components
  - Buttons
  - Forms
  - Icons
- Mobile first: highly responsive layout

Javascript?

- Interpreted language brought into the browser
- Not really related to Java in any way - formally ECMAScript
- Why?
  - HTML is not a programming language
  - CSS is not a programming language (well, ...)
- Would still like to have “programmability” inside browser
- Not part of the core presentation requirements
  - Very useful, but will be considered later

Summary

- Presentation - Human interaction
- Separate content from style
  - Markup - HTML
  - Styling - CSS