# The TEX hyphenation applied to HTML About Frank M. Liangs hyphenation algorithm and its

About Frank M. Liangs hyphenation algorithm and it port to Javascript

Mathias Nater mathias@mnn.ch http://mnn.ch/

BachoTFX 2010

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm [1983]

Creating the patterns (patgen)

Jsing the patterns hyphenation)

HTML and the soft hyphen

The Port to

erver side or Client si

ow it works

mprovements

### Motivation

Text layout without hyphenation Text layout with hyphenation

#### Hyphenation for HTML



### Motivation

Text layout without hyphenation Text layout with hyphenation

### The TFX hyphenation algorithm

The original TFX hyphenation algorithm (1977) The current TFX hyphenation algorithm (1983) Creating the patterns (patgen) Using the patterns (hyphenation)

#### Hyphenation for HTML



### Motivation

Text layout without hyphenation Text layout with hyphenation

### The TFX hyphenation algorithm

The original TFX hyphenation algorithm (1977) The current TFX hyphenation algorithm (1983) Creating the patterns (patgen) Using the patterns (hyphenation)

### HTML and the soft hyphen

#### Hyphenation for HTML

### Motivation

Text layout without hyphenation Text layout with hyphenation

### The TFX hyphenation algorithm

The original TFX hyphenation algorithm (1977) The current TFX hyphenation algorithm (1983) Creating the patterns (patgen) Using the patterns (hyphenation)

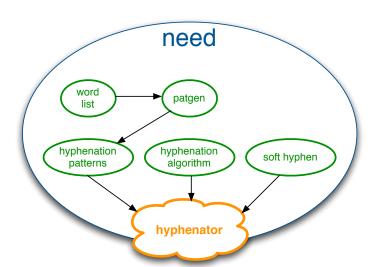
### HTML and the soft hyphen

### The Port to Javascript

Server side or Client side? How it works Differences and Improvements Back to the Future

#### Hyphenation for HTML





#### Hyphenation for **HTML**

### Current Browsers

- ▶ MS IE 6/7/8 ( $\sim 44\%$ )
- ► Firefox 3.5 (~ 42%)
- ► Safari 4 (~ 4%)
- ► Opera 10 (~ 3%)

### do not hyphenate text automatically!

- ▶ align left: overfull boxes and unbalanced line endings
- justified: big word spaces and rivers

### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

#### layout $\ensuremath{\text{w}/\text{o}}$ hyphenation

layout with hyphena

The TEX hyphenation algorithm

The original TEX hyphenation algorithm

he current TEX yphenation algorith 983)

Creating the patterns (patgen)

Jsing the patterns

HTML and the

## The Port to

Server side or Client side How it works

Darline she France

### Current Browsers

- ► MS IE 6/7/8 (~ 44%)
- ► Firefox 3.5 (~ 42%)
- ► Safari 4 (~ 4%)
- ► Opera 10 (~ 3%)

### do not hyphenate text automatically!

- ▶ align left: overfull boxes and unbalanced line endings
- justified: big word spaces and rivers

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

### layout w/o hyphenation

layout with hyphenation

The TEX hyphenation

The original TEX

1977) The current TEX

ypnenation algori 1983) Creating the patti

(patgen)

Jsing the patterns hyphenation)

HTML and the

The Port to

The Port to Javascript

Server side or Client sid How it works

Differences and Improvements

### Current Browsers

- ▶ MS IE 6/7/8 ( $\sim 44\%$ )
- ► Firefox 3.5 (~ 42%)
- ▶ Safari 4 (~ 4%)
- Opera 10 ( $\sim 3\%$ )

### do not hyphenate text automatically! This leads to poor typography:

- align left: overfull boxes and unbalanced line endings
- justified: big word spaces and rivers

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

#### layout w/o hyphenation

### Current Browsers

- ▶ MS IE 6/7/8 ( $\sim 44\%$ )
- ▶ Firefox 3.5 ( $\sim$  42%)
- ▶ Safari 4 (~ 4%)
- ▶ Opera 10 (~ 3%)

do not hyphenate text automatically!

This leads to poor typography:

- align left: overfull boxes and unbalanced line endings
- justified: big word spaces and rivers

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

layout with hyphenati

The TEX

The original TEX hyphenation algorithm

he current TEX phenation algorithr

Creating the pattern (patgen)

Using the patterns (hyphenation)

HTML and the

soft hyphen

The Port to Javascript

> Server side or Client side How it works

Differences and

### Current Browsers

- ▶ MS IE 6/7/8 ( $\sim 44\%$ )
- ► Firefox 3.5 (~ 42%)
- ▶ Safari 4 (~ 4%)
- ▶ Opera 10 (~ 3%)

do not hyphenate text automatically!

This leads to poor typography:

- align left: overfull boxes and unbalanced line endings
- justified: big word spaces and rivers

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivatio

layout w/o hyphenation

The T<sub>E</sub>X

hyphenation algorithm

The original TEX hyphenation algorith (1977)

> he current TEX rphenation algorithr 983)

Creating the patterns (patgen)

Using the patterns

ITML and the

soft hyphen

The Port to Javascript

Server side or Client sid How it works

How it works Differences and



text-align: left;

DAS	S N	<b>1</b> A	G	4ZI	N
2010/4	Letztes	Heft I	Archiv I	Nächstes	Heft

DAS MAGAZIN





### DER MÄDCHEN-EFFEKT

von Birgit Schmid, 1 Kommentar

Afrika braucht mehr Frauen wie Edna Adan, Mit ihrem Spital in Somaliland zeigt sie, wie man Extremismus und Armut bekämpfen kann.

#### WENN HELVETIA HADERT

von Peter Ziegler, 8 Kommentare

In diesen wirren Zeiten dominieren die Kopfjäger, Alleingänger und Ouick-Fix-Technokraten die innenpolitische Diskussion. Doch was es wirklich braucht. ist eine Regierungsreform, eine gescheite Bankenregulierung, und, ja, den EU-Beitritt.

#### PROJEKT 80\*81

von Das Magazin Die Achtzigerjahre in Bildern ALIS DER REDAKTION Projekt 80\*81 IOURNAL DER GEGENWART

Verantwortungsscheue Volksvertreter Glücklich werden #4

Plüss, empirisch

KÜNG

Max Küng MONATSCESPRÄCH

Alain de Botton Teil 4 HIER FRAGT DER CHEF Finn Canonica auf DRS 4 Talk

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

layout w/o hyphenation

text-align: justify;

Anmelden | Mein Profil

Suchen

## DAS MAGAZIN



#### DER MÄDCHEN-EFFEKT

von Rirgit Schmid. 1 Kommentar

Afrika braucht mehr Frauen wie Edna Adan, Mit ihrem Spital in Somaliland zeigt sie, wie man Extremismus und Armut bekämpfen kann.

#### WENN HELVETIA HADERT

von Peter Ziegler, 8 Kommentare

In diesen wirren Zeiten dominieren die Kopfjäger, Alleingänger und Quick-Fix-Technokraten innenpolitische Diskussion. Doch was es wirklich braucht, Regierungsreform, eine gescheite Bankenregulierung, und, ja, den EU-Beitritt.

#### PROIEKT 80\*81

von Das Magazin Die Achtzigeriahre in Bildern AUS DER REDAKTION Projekt 80\*81

**IOURNAL DER GEGENWART** 

Verantwortungsscheue Volksvertreter Glücklich werden #4

Plüss, empirisch

KÜNG Max Küng

MONATSGESPRÄCH

Alain de Botton Teil 4 HIER FRAGT DER CHEF Finn Canonica auf DRS 4 Talk

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

layout w/o hyphenation

text-align: left;

DAS MAGAZIN









von Birgit Schmid, 1 Kommentar

Afrika braucht mehr Frauen wie Edna Adan, Mit ihrem Spital in Somaliland zeigt sie, wie man Extremismus und Armut bekämpfen kann.

#### WENN HELVETIA HADERT

von Peter Ziegler, 8 Kommentare

In diesen wirren Zeiten dominieren die Kopfjäger, Alleingänger und Ouick-Fix-Technokraten die innenpolitische Diskussion. Doch was es wirklich braucht. ist eine Regierungsreform, eine gescheite Bankenregulierung, und, ja, den EU-Beitritt.

#### PROJEKT 80\*81

von Das Magazin

Die Achtzigerjahre in Bildern

ALIS DER REDAKTION Projekt 80\*81

IOURNAL DER GEGENWART Verantwortungsscheue Volksvertreter

Glücklich werden #4

Plüss, empirisch

KÜNG Max Küng

MONATSCESPRÄCH

Alain de Botton Teil 4 HIER FRAGT DER CHEF Finn Canonica auf DRS 4 Talk

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

layout with hyphenation

text-align: left;

Anmelden | Mein Profil

Suchen

## DAS MAGAZIN



#### DER MÄDCHEN-EFFEKT

von Birgit Schmid, 1 Kommentar

Afrika braucht mehr Frauen wie Edna Adan. Mit ihrem Spital in Somaliland zeigt sie, wie man Extremismus und Armut bekämpfen kann.

#### WENN HELVETIA HADERT

von Peter Ziegler, 8 Kommentare

In diesen wirren Zeiten dominieren die Kopfjäger, Alleingänger und Quick-Fix-Technokraten die innenpolitische Diskussion. Doch was es wirklich braucht, ist eine Regierungsreform, eine gescheite Bankenregulierung, und, ja, den RII-Reitritt

#### PROJEKT 80\*81

von Das Magazin

Die Achtzigerjahre in Bildern

AUS DER REDAKTION Projekt 80\*81 JOURNAL DER GEGENWART

Verantwortungsscheue Volksvertreter Glücklich werden #4

Plüss, empirisch KÜNG

Max Küng

MONATSGESPRÄCH Alain de Botton Teil 4

Alain de Botton Teil 4
HIER FRAGT DER CHEF
Finn Canonica auf DRS 4 Talk

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

iviotivation

layout w/o hyphenation layout with hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

Creating the pattern (patgen)

Using the patterns (hyphenation)

HTML and the

The Port to

Server side or Client side?

How it works

Darline she Firein

◆ロト 4個ト 4 豆ト 4 豆ト 豆 目 め 9 (ペ)

text-align: justify;

Anmelden | Mein Profil

Suchen

## DAS MAGAZIN



#### DER MÄDCHEN-EFFEKT

von Rirgit Schmid. 1 Kommentar

Afrika braucht mehr Frauen wie Edna Adan, Mit ihrem Spital in Somaliland zeigt sie, wie man Extremismus und Armut bekämpfen kann.

#### WENN HELVETIA HADERT

von Peter Ziegler, 8 Kommentare

In diesen wirren Zeiten dominieren die Kopfjäger, Alleingänger und Quick-Fix-Technokraten innenpolitische Diskussion. Doch was es wirklich braucht, Regierungsreform, eine gescheite Bankenregulierung, und, ja, den EU-Beitritt.

#### PROIEKT 80\*81

von Das Magazin Die Achtzigeriahre in Bildern AUS DER REDAKTION Projekt 80\*81

**IOURNAL DER GEGENWART** Verantwortungsscheue Volksvertreter

Glücklich werden #4

Plüss, empirisch KÜNG

Max Küng

MONATSGESPRÄCH

Alain de Botton Teil 4 HIER FRAGT DER CHEF Finn Canonica auf DRS 4 Talk

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

layout with hyphenation

text-align: justify;

Anmelden | Mein Profil

Suchen

### DAS MAGAZIN



#### DER MÄDCHEN-EFFEKT

von Birgit Schmid. 1 Kommentar

Afrika braucht mehr Frauen wie Edna Adan. Mit ihrem Spital in Somaliland zeigt sie, wie man Extremismus und Armut bekämpfen kann.

#### WENN HELVETIA HADERT

von Peter Ziegler, 8 Kommentare

In diesen wirren Zeiten dominieren die Kopfjäger, Alleingänger und Quick-Fix-Technokraten die innenpolitische Diskussion. Doch was es wirklich braucht, ist eine Regierungsreform, eine gescheite Bankenregulierung, und, ja, den EU-Beitritt.

#### PROJEKT 80\*81

von Das Magazin

Die Achtzigerjahre in Bildern

AUS DER REDAKTION Projekt 80\*81

JOURNAL DER GEGENWART

Verantwortungsscheue Volksvertreter Glücklich werden #4

Plüss, empirisch

KÜNG Max Küng

MONATSGESPRÄCH Alain de Botton Teil 4

Alain de Botton Teil 4 HIER FRAGT DER CHEF Finn Canonica auf DRS 4 Talk

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Viotivation

layout w/o hyphenation

The T<sub>E</sub>X

The original TEX hyphenation algorithm

(1977) The current TEX hyphenation algorithm

1983) Creating the patterns

(patgen) Using the patterns

HTML and the soft hyphen

The Port to

Server side or Client side?

low it works

Saala ka kha Eukona

# We need (automatic) hyphenation in HTML!

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivatior

layout w/o hyphenation layout with hyphenation

hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorith (1983)

Creating the patter (patgen)

Using the patterns

HTML and the

he Port to

avascript erver side or Client s

ow it works

Rock to the Euture

## The original TEX hyphenation algorithm

The original hyphenation algorithm

### 1977 by Donald E. Knuth and Franklin M. Liang

- for english only
- suffix and prefix removal

- small exception dictionary

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

## The original TFX hyphenation algorithm

The original hyphenation algorithm

### 1977 by Donald E. Knuth and Franklin M. Liang

- for english only
- suffix and prefix removal
- vowel-consonant-consonant-vowel breaking
- special case rules (e.g. "break after ck!")
- ► small exception dictionary

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

1983) Creating the patterns

(patgen) Using the pattern:

TML and the

HTML and the soft hyphen

The Port to Javascript

Server side or Client sid How it works

Differences and Improvements

## The original TEX hyphenation algorithm

The original hyphenation algorithm

### 1977 by Donald E. Knuth and Franklin M. Liang

- for english only
- suffix and prefix removal
- special case rules (e.g. "break after ck!")
- small exception dictionary

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

The original TFX hyphenation algorithm

(1977)

## The original T<sub>F</sub>X hyphenation algorithm

The original hyphenation algorithm

### 1977 by Donald E. Knuth and Franklin M. Liang

- for english only
- suffix and prefix removal
- vowel-consonant-consonant-vowel breaking
- special case rules (e.g. "break after ck!")
- small exception dictionary

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

## The original T<sub>F</sub>X hyphenation algorithm

The original hyphenation algorithm

### 1977 by Donald E. Knuth and Franklin M. Liang

- for english only
- suffix and prefix removal
- vowel-consonant-consonant-vowel breaking
- special case rules (e.g. "break after ck!")

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

## The original T<sub>F</sub>X hyphenation algorithm

The original hyphenation algorithm

### 1977 by Donald E. Knuth and Franklin M. Liang

- for english only
- suffix and prefix removal
- vowel-consonant-consonant-vowel breaking
- special case rules (e.g. "break after ck!")
- small exception dictionary

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

## The original TFX hyphenation algorithm

The original hyphenation algorithm

### 1977 by Donald E. Knuth and Franklin M. Liang

- for english only
- suffix and prefix removal
- vowel-consonant-consonant-vowel breaking
- special case rules (e.g. "break after ck!")
- small exception dictionary

Found  $\sim 40\%$  of the allowable hyphen points with 1% error

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

viotivation

layout w/o hyphenation layout with hyphenation

The TEX

Typhenation

Algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

1983) Creating the patterns

(patgen) Using the patterns

(hyphenation)

HTML and the soft hyphen

he Port to

avascript

How it works

Differences and

## The current T<sub>F</sub>X hyphenation algorithm

The current hyphenation algorithm

### 1983 PhD thesis by Franklin M. Liang

- use of hyphenation patterns
- ▶ two algorithms:
  - pattern creation (patgen)
  - applying the patterns (TEX)
- support for a wide range of languages
- ► small, easy, fast

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivatio

layout w/o hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm

The current TEX hyphenation algorithm (1983)

Creating the patterns (patgen)

patgen) Jsing the patterns

ITML and the

soft hyphen

The Port to Javascript

> Server side or Client side How it works

Darline she Franci

## The current T<sub>F</sub>X hyphenation algorithm

The current hyphenation algorithm

### 1983 PhD thesis by Franklin M. Liang

- use of hyphenation patterns
- two algorithms:
  - pattern creation (patgen)
  - applying the patterns (TEX)
- support for a wide range of languages
- small, easy, fast

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm

The current TEX hyphenation algorithm (1983)

Creating the patterns (patgen)

Jsing the patterns hyphenation)

HTML and the

The Port to

Server side or Client side How it works

Back to the Euture

イロト 4周ト 4 ヨト 4 ヨト ヨ|モ めなべ

## The current T<sub>F</sub>X hyphenation algorithm

The current hyphenation algorithm

### 1983 PhD thesis by Franklin M. Liang

- use of hyphenation patterns
- two algorithms:
  - pattern creation (patgen)
  - applying the patterns (TFX)
- support for a wide range of languages
- small, easy, fast

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

The current TFX hyphenation algorithm (1983)



## The current TFX hyphenation algorithm

The current hyphenation algorithm

### 1983 PhD thesis by Franklin M. Liang

- use of hyphenation patterns
- ▶ two algorithms:
  - pattern creation (patgen)
  - ▶ applying the patterns (TEX)
- support for a wide range of languages
- ► small, easy, fast

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The T<sub>E</sub>X hyphenation algorithm

The original TEX hyphenation algorithm

The current TEX hyphenation algorithm (1983)

Creating the patterns (patgen)

Jsing the patterns

HTML and th

The Port to

Server side or Client side? How it works

Differences and



# ► INPUT: a list of hyphenated words [, precomputed pattern, translate file]

- takes up to 9 runs (asking for many settings, adding a new level in each run)
- ► OUTPUT: pattern file, statistics (a lot!)
- old code
- ▶ no UTF-8
- refactored by David Antoš (OPatGen), but doesn't compile

### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation

### he T<sub>E</sub>X yphenation

The original TEX hyphenation algorithm

1977) The current TEX syphenation algorithm

yphenation algorithm 1983)

### Creating the patterns (patgen)

hyphenation)

HTML and the soft hyphen

## The Port to

Server side or Client side? How it works

### ▶ INPUT: a list of hyphenated words [, precomputed pattern, translate file

- ▶ takes up to 9 runs (asking for many settings, adding a
- ▶ old code
- ▶ no UTF-8
- refactored by David Antoš (OPatGen), but doesn't

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Creating the patterns

### (patgen)



- ▶ INPUT: a list of hyphenated words [, precomputed pattern, translate file
- takes up to 9 runs (asking for many settings, adding a new level in each run)
- old code
- ▶ no UTF-8
- refactored by David Antoš (OPatGen), but doesn't

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/



- ▶ INPUT: a list of hyphenated words [, precomputed pattern, translate file
- takes up to 9 runs (asking for many settings, adding a new level in each run)
- OUTPUT: pattern file, statistics (a lot!)
- old code
- ▶ no UTF-8
- refactored by David Antoš (OPatGen), but doesn't

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/



- ▶ INPUT: a list of hyphenated words [, precomputed pattern, translate file
- takes up to 9 runs (asking for many settings, adding a new level in each run)
- OUTPUT: pattern file, statistics (a lot!)
- old code
- ▶ no UTF-8
- refactored by David Antoš (OPatGen), but doesn't

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/



- ▶ INPUT: a list of hyphenated words [, precomputed pattern, translate file
- takes up to 9 runs (asking for many settings, adding a new level in each run)
- OUTPUT: pattern file, statistics (a lot!)
- old code
- ▶ no UTF-8
- refactored by David Antoš (OPatGen), but doesn't compile

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

## applying the patterns

- in1 h212 4edi h41e
- patterns: short strings with integer values
- odd values: valid breakpoints
- even values: forbidden breakpoints
- lower values are overwritten by higher values
- points mark begin/end of the word

### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation

#### The T<sub>E</sub>X hyphenation

algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm (1983)

Creating the patterns (patgen)

Using the patterns (hyphenation)

TML and the

### soft hyphen

## The Port to Javascript

Server side or Client sid How it works

improvements

- ▶ .in1 b2l2 4edi b4le.
- patterns: short strings with integer values
- odd values: valid breakpoints
- even values: forbidden breakpoints
- lower values are overwritten by higher values
- points mark begin/end of the word

### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

The TEX hyphenation

The original TEX hyphenation algorithm

(1977)
The current TEX
hyphenation algorithm

(1983)
Creating the patt

(patgen) Using the patterns

(hyphenation)

HTML and the soft hyphen

The Port to Javascript

Server side or Client side How it works

Deal, se she France



- ▶ .in1 b2l2 4edi b4le.
- patterns: short strings with integer values
- odd values: valid breakpoints
- even values: forbidden breakpoints
- ▶ lower values are overwritten by higher values
- ▶ points mark begin/end of the word

### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1077)

The current TEX hyphenation algorithm

Creating the pattern (patgen)

Using the patterns (hyphenation)

HTML and the

soft hyphen

The Port to Javascript

Server side or Client si How it works

Differences and

- .in1 b212 4edi b41e.
- patterns: short strings with integer values
- odd values: valid breakpoints
- even values: forbidden breakpoints
- lower values are overwritten by higher values
- points mark begin/end of the word

#### Hyphenation for HTML

### Mathias Nater mathias@mnn.ch http://mnn.ch/

Using the patterns (hyphenation)



- .in1 b212 4edi b41e.
- patterns: short strings with integer values
- odd values: valid breakpoints
- even values: forbidden breakpoints
- lower values are overwritten by higher values
- points mark begin/end of the word

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Using the patterns (hyphenation)



### incredible

. i n5c2r4e d3i1b4l2e

### Hyphenation for HTML

### Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation

# The T<sub>E</sub>X hyphenation

hyphenation algorithm

The original TEX hyphenation algorithm (1977)

> he current TEX yphenation algorith 1983)

reating the patter patgen)

Using the patterns (hyphenation)

HTML and the

### The Port to Javascript

Server side or Client side. How it works

Improvements

```
incredible
.incredible.
```

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Using the patterns (hyphenation)

```
incredible
.incredible.
. i n1
               b212
         4e d i
             i1b 1
    n1c r
               b41 e.
     5cred
          e d3i b
       2r2e d
```

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Using the patterns (hyphenation)

◆□▶ ◆周▶ ◆臣▶ ◆臣▶ 重1章 釣Q@

```
incredible
.incredible.
. i n1
                b212
         4e d i
              i1b 1
    n1c r
                b41 e.
     5cred
          e d3i b
       2r2e d
 . i n5c2r4e d3i1b4l2e .
```

Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Using the patterns (hyphenation)

◆□▶ ◆周▶ ◆臣▶ ◆臣▶ 重1章 釣Q@

```
incredible
.incredible.
. i n1
                b212
         4e d i
              i1b 1
    n1c r
                b41 e .
     5cred
          e d3i b
       2r2e d
  i n5c2r4e d3i1b4l2e .
in-cred-i-ble
```

Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The T<sub>E</sub>X hyphenation algorithm

The original TEX hyphenation algorithm

he current TEX

Creating the patter (natgen)

Using the patterns (hyphenation)

TML and the

The Port to Javascript

Server side or Client sid How it works

Differences and

Back to the Future

example 2

- ▶ limited control over textlayout (text-align: left | right | justify
- manual line breaks (<br>)
- manually inserted soft hyphens
  (­ discretionary hyphen)
- some more controls are upcoming with CSS3
- laying out text is up to the browser
- developer has no control over how text is displayed

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation

# The TEX hyphenation

algorithm
The original TEX

hyphenation algorithm (1977)

he current TEX yphenation algorithm 1983)

Creating the patterns patgen)

hyphenation)

## HTML and the soft hyphen

# The Port to Javascript

Server side or Client side How it works

- limited control over textlayout
   (text-align: left | right | justify)
- manual line breaks (<br>)
- manually inserted soft hyphens
  (­ discretionary hyphen)
- some more controls are upcoming with CSS3
- laying out text is up to the browser
- developer has no control over how text is displayed

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation

# The TEX hyphenation

algorithm
The original TEX

hyphenation algorithm (1977)

hyphenation algorithm 1983)

reating the patterns patgen) Jsing the patterns

### HTML and the

# soft hyphen

# The Port to Javascript

Server side or Client sid How it works

- limited control over textlayout
   (text-align: left | right | justify)
- manual line breaks (<br>)
- manually inserted soft hyphens(­ discretionary hyphen)
- some more controls are upcoming with CSS3
- laying out text is up to the browser
- developer has no control over how text is displayed

### Hyphenation for HTML

### Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation

### The T<sub>E</sub>X hyphenation

algorithm

hyphenation algorithm (1977)

ne current 1EX yphenation algorith 1983)

Creating the pattern: (patgen) Using the patterns

Using the patterns (hyphenation)

### HTML and the soft hyphen

# The Port to

Server side or Client sid

low it works

Improvements



- limited control over textlayout (text-align: left | right | justify)
- manual line breaks (<br>)
- manually inserted soft hyphens (­ - discretionary hyphen)
- some more controls are upcoming with CSS3
- developer has no control over how text is displayed

#### Hyphenation for HTML

### Mathias Nater mathias@mnn.ch http://mnn.ch/

#### HTML and the soft hyphen



- limited control over textlayout
   (text-align: left | right | justify)
- manual line breaks (<br>)
- manually inserted soft hyphens (­ - discretionary hyphen)
- some more controls are upcoming with CSS3
- laying out text is up to the browser
- developer has no control over how text is displayed

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

The T<sub>E</sub>X hyphenation algorithm

The original TEX hyphenation algorithm

he current TEX yphenation algorith (983)

Creating the patterns (patgen)

Using the patterns 'hyphenation)

# HTML and the soft hyphen

The Port to Javascript

Server side or Client sid

low it works



### Text layout with hyphenation

text-align: justify;

Anmelden | Mein Profil

Suchen

# DAS MAGAZIN



#### DER MÄDCHEN-EFFEKT

von Rirgit Schmid. 1 Kommentar

Afrika braucht mehr Frauen wie Edna Adan, Mit ihrem Spital in Somaliland zeigt sie, wie man Extremismus und Armut bekämpfen kann.

#### WENN HELVETIA HADERT

von Peter Ziegler, 8 Kommentare

In diesen wirren Zeiten dominieren die Kopfjäger, Alleingänger und Quick-Fix-Technokraten innenpolitische Diskussion. Doch was es wirklich braucht, Regierungsreform, eine gescheite Bankenregulierung, und, ja, den EU-Beitritt.

#### PROIEKT 80\*81

von Das Magazin

Die Achtzigeriahre in Bildern

AUS DER REDAKTION

Projekt 80\*81 **IOURNAL DER GEGENWART** 

Verantwortungsscheue Volksvertreter Glücklich werden #4

Plüss, empirisch KÜNG

Max Küng

MONATSGESPRÄCH

Alain de Botton Teil 4 HIER FRAGT DER CHEF Finn Canonica auf DRS 4 Talk

#### Hyphenation for HTML

### Mathias Nater mathias@mnn.ch http://mnn.ch/

#### HTML and the soft hyphen

### Text layout with hyphenation

text-align: justify;

Anmelden | Mein Profil

Suchen

### DAS MAGAZIN



#### DER MÄDCHEN-EFFEKT

von Birgit Schmid. 1 Kommentar

Afrika braucht mehr Frauen wie Edna Adan. Mit ihrem Spital in Somaliland zeigt sie, wie man Extremismus und Armut bekämpfen kann.

#### WENN HELVETIA HADERT

von Peter Ziegler, 8 Kommentare

In diesen wirren Zeiten dominieren die Kopfjäger, Alleingänger und Quick-Fix-Technokraten die innenpolitische Diskussion. Doch was es wirklich braucht, ist eine Regierungsreform, eine gescheite Bankenregulierung, und, ja, den EII-Beitritt.

#### PROJEKT 80\*81

von Das Magazin

Die Achtzigeriahre in Bildern

AUS DER REDAKTION Projekt 80\*81

JOURNAL DER GEGENWART

Verantwortungsscheue Volksvertreter Glücklich werden #4

Plüss, empirisch

KÜNG

Max Küng

MONATSGESPRÄCH
Alain de Botton Teil 4
HIER FRAGT DER CHEF
Finn Canonica auf DRS 4 Talk

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

iviotivation

layout w/o hyphenation layout with hyphenation

The T<sub>E</sub>X hyphenation

The original TEX hyphenation algorithm

hyphenation algorithm (1977) The current TEX

yphenation algorithm 1983)

(patgen)

Jsing the patterns hyphenation)

### HTML and the soft hyphen

The Port to

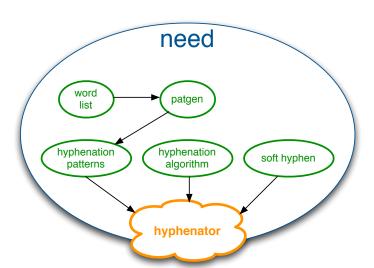
Server side or Client side

low it works

- - -



# Putting all together



#### Hyphenation for **HTML**

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

### The Port to Javascript

### Pro server side:

- lower bandwidth usage
- faster
- only hyphenate once, store the result

### Pro client side:

- cleaner HTML (search engines!)
- takes in count client oddities
- can be switched on/off
- ▶ hyphenation is part of CSS3, so even the W3C believes that hyphenation belongs to the client
- user generated text can be hyphenated on the fly

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation

The TEX hyphenation

The original TEX

hyphenation algorithm (1977)

"he current TEX hyphenation algorithm 1983)

Creating the patterns patgen)

nyphenation)

HTML and the soft hyphen

The Port to

#### Server side or Client side?

How it works Differences and

### Pro server side:

- lower bandwidth usage
- faster
- only hyphenate once, store the result

### Pro client side:

- cleaner HTML (search engines!)
- takes in count client oddities
- can be switched on/off
- hyphenation is part of CSS3, so even the W3C believes that hyphenation belongs to the client
- user generated text can be hyphenated on the fly

### Hyphenation for HTML

### Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation

The T<sub>E</sub>X hyphenation

The original TEX

hyphenation algorithm (1977)

he current TEX yphenation algorith 1983)

creating the patterns patgen)

yphenation)

HTML and the soft hyphen

he Port to

#### Server side or Client side?

How it works Differences and



### Pro server side:

- lower bandwidth usage
- faster
- only hyphenate once, store the result

### Pro client side:

- cleaner HTML (search engines!)
- takes in count client oddities
- can be switched on/off
- hyphenation is part of CSS3, so even the W3C believes

#### Hyphenation for HTML

### Mathias Nater mathias@mnn.ch http://mnn.ch/

### Server side or Client side?



### Pro server side:

- lower bandwidth usage
- faster
- only hyphenate once, store the result

### Pro client side:

- cleaner HTML (search engines!)
- ▶ takes in count client oddities
- can be switched on/off
- hyphenation is part of CSS3, so even the W3C believes that hyphenation belongs to the client
- user generated text can be hyphenated on the fly

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

iviotivation

layout w/o hyphenation layout with hyphenation

The TEX hyphenation

The original TEX hyphenation algorithm

(1977) The current TEX hyphenation algorithm

1983) Creating the pattern

patgen) Using the patterns

TML and the

soft hyphen

he Port to

Server side or Client side?

low it works



### Pro server side:

- lower bandwidth usage
- faster
- only hyphenate once, store the result

### Pro client side:

- cleaner HTML (search engines!)
- takes in count client oddities
- can be switched on/off
- hyphenation is part of CSS3, so even the W3C believes that hyphenation belongs to the client
- user generated text can be hyphenated on the fly

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

iviotivation

layout w/o hyphenation

The TEX
hyphenation

The original TEX hyphenation algorithm

1977) The current TEX hyphenation algorithm

1983) Creating the patter

(patgen) Using the patterns

TML and the

H I ML and the soft hyphen

he Port to

Server side or Client side?

How it works Differences and



### Pro server side:

- lower bandwidth usage
- faster
- only hyphenate once, store the result

### Pro client side:

- cleaner HTML (search engines!)
- takes in count client oddities
- can be switched on/off
- hyphenation is part of CSS3, so even the W3C believes that hyphenation belongs to the client
- user generated text can be hyphenated on the fly

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

iviotivation

layout w/o hyphenation layout with hyphenation

The TEX hyphenation

The original TEX hyphenation algorithm

(1977)
The current TEX
hyphenation algorithm

1983) Creating the patter

(patgen) Using the patterns

hyphenation)

HTML and the soft hyphen

The Port to

Server side or Client side?

How it works



### Pro server side:

- lower bandwidth usage
- faster
- only hyphenate once, store the result

### Pro client side:

- cleaner HTML (search engines!)
- takes in count client oddities
- can be switched on/off
- hyphenation is part of CSS3, so even the W3C believes that hyphenation belongs to the client
- user generated text can be hyphenated on the fly

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

iviotivation

layout w/o hyphenation

The TEX

Typhenation

The original TEX hyphenation algorithm (1977)

The current TEX

Creating the patterns (patgen)

Using the patterns (hyphenation)

HTML and the soft hyphen

he Port to

Server side or Client side?

Server side or Client side?

How it works
Differences and



### Pro server side:

- lower bandwidth usage
- faster
- only hyphenate once, store the result

### Pro client side:

- cleaner HTML (search engines!)
- takes in count client oddities
- can be switched on/off
- hyphenation is part of CSS3, so even the W3C believes that hyphenation belongs to the client
- user generated text can be hyphenated on the fly

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

he T<sub>E</sub>X yphenation Igorithm

The original TEX hyphenation algorithm (1977)

The current TEX
hyphenation algorithm

Creating the pattern (patgen)

Jsing the patterns

TML and the

he Port to

Javascript

Server side or Client side?

How it works Differences and



### Pro server side:

- lower bandwidth usage
- faster
- only hyphenate once, store the result

### Pro client side:

- cleaner HTML (search engines!)
- takes in count client oddities
- can be switched on/off
- hyphenation is part of CSS3, so even the W3C believes that hyphenation belongs to the client
- user generated text can be hyphenated on the fly

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

> he T<sub>E</sub>X yphenation Igorithm

The original TEX hyphenation algorithm (1977)

> he current TEX yphenation algorithm 1983)

Creating the pattern (patgen)

Jsing the patterns hyphenation)

HTML and the oft hyphen

The Port to

Server side or Client side?

How it works Differences and

talian da Cia



- server side solutions already existed: php, perl, java, python
- ▶ I believe that hyphenation has to be done in the clien
- Javascript is a very interesting language
- the acceptance of Javascript is growing
- ► Firefox 2 didn't support ­
- ▶ I like bookmarklets
- hyphenator.js: client-side hyphenation
- ▶ it's proofing to be a good decision:
  - other webkit based programs are using hyphenator
  - it's easy to use
  - there's a big effort on making javascript faster

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

Creating the patterns patgen)

yphenation)

HTML and the soft hyphen

The Port to

Server side or Client side?

low it works



- server side solutions already existed: php, perl, java, python
- I believe that hyphenation has to be done in the client

- ▶ I like bookmarklets

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Server side or Client side?



- server side solutions already existed: php, perl, java, python
- I believe that hyphenation has to be done in the client
- Javascript is a very interesting language
- the acceptance of Javascript is growing
- ► Firefox 2 didn't support ­
- ▶ I like bookmarklets
- hyphenator.js: client-side hyphenation
- ▶ it's proofing to be a good decision:
  - other webkit based programs are using hyphenator
    - ▶ it's easy to use
  - there's a big effort on making javascript faste

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

he current TEX hyphenation algorithm 1983)

Creating the patterns (patgen)

sing the patterns

TML and the oft hyphen

The Port to

Server side or Client side?

low it works

provements



- server side solutions already existed: php, perl, java, python
- I believe that hyphenation has to be done in the client
- Javascript is a very interesting language
- the acceptance of Javascript is growing
- ► Firefox 2 didn't support ­
- ▶ I like bookmarklets
- hyphenator.js: client-side hyphenation
- ▶ it's proofing to be a good decision:
  - other webkit based programs are using hyphenator
  - ▶ it's easy to use
  - there's a big effort on making javascript faster

### Hyphenation for HTML

### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

> The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

he current TEX hyphenation algorithm 1983)

Creating the patterns (patgen)

sing the patterns yphenation)

HTML and the soft hyphen

The Port to

Server side or Client side?

Server side or Client side?

low it works Differences and



- server side solutions already existed: php, perl, java, python
- ▶ I believe that hyphenation has to be done in the client
- Javascript is a very interesting language
- the acceptance of Javascript is growing
- Firefox 2 didn't support ­
- ▶ I like bookmarklets
- hyphenator.js: client-side hyphenation
- ▶ it's proofing to be a good decision:
  - other webkit based programs are using hyphenator
  - ▶ it's easy to use
  - there's a big effort on making javascript faster

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

> he T<sub>E</sub>X yphenation lgorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm 1983)

Creating the patterns patgen)

ing the patterns yphenation)

HTML and the soft hyphen

The Port to

Server side or Client side?

low it works

provements



- server side solutions already existed: php, perl, java, python
- ▶ I believe that hyphenation has to be done in the client
- Javascript is a very interesting language
- the acceptance of Javascript is growing
- Firefox 2 didn't support ­
- ► I like bookmarklets
- hyphenator.js: client-side hyphenation
- ▶ it's proofing to be a good decision:
  - other webkit based programs are using hyphenator
  - it's easy to use
  - there's a big effort on making javascript faster

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

iviotivation

layout w/o hyphenation layout with hyphenation

he T<sub>E</sub>X syphenation slgorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm 1983)

Creating the patterns patgen)

sing the patterns yphenation)

HTML and the soft hyphen

The Port to

Server side or Client side?

low it works

Oifferences and



- server side solutions already existed: php, perl, java, python
- ▶ I believe that hyphenation has to be done in the client
- Javascript is a very interesting language
- the acceptance of Javascript is growing
- Firefox 2 didn't support ­
- ► I like bookmarklets
- hyphenator.js: client-side hyphenation
- ▶ it's proofing to be a good decision:
  - ▶ other webkit based programs are using hyphenator
  - it's easy to use
  - there's a big effort on making javascript faster

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

he T<sub>E</sub>X yphenation lgorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm (1983)

Creating the patterns patgen)

sing the patterns yphenation)

HTML and the soft hyphen

The Port to

Server side or Client side?

ow it works



- server side solutions already existed: php, perl, java, python
- ▶ I believe that hyphenation has to be done in the client
- Javascript is a very interesting language
- the acceptance of Javascript is growing
- Firefox 2 didn't support ­
- ▶ I like bookmarklets
- hyphenator.js: client-side hyphenation
- ▶ it's proofing to be a good decision:
  - other webkit based programs are using hyphenator
  - it's easy to use
  - there's a big effort on making javascript faster

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

he TEX phenation gorithm

The original TEX hyphenation algorithm (1977)

he current TEX hyphenation algorithm 1983)

Creating the patterns patgen)

sing the patterns yphenation)

HTML and the soft hyphen

The Port to

Server side or Client side?

Server side or Client side How it works

How it works Differences and



### How it works

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

### How it works

- execution is fast.
- downloading the script and the patterns takes time

### How it works

### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

### 1. register all elements that need hyphenation

- if the language is not set, ask for it
- 3. download the patterns, if not already done
- 4. split the paragraphs in words (and URLs)
- 5. process each word, put ­ at every valid breakpoint
- 6. The browser will re-render the text automatically, taking in account the soft hyphens.
- execution is fast.
- ▶ downloading the script and the patterns takes time script: 25 KB, en: 25 KB, pl: 37 KB, de: 74 KB

#### Motivation

layout w/o hyphenation layout with hyphenation

### he TEX yphenation

The original TEX hyphenation algorithm

The current TEX hyphenation algorithm

1983) Creating the patterns (patgen)

Jsing the patterns hyphenation)

HTML and the

## The Port to

Server side or Client side?

### How it works

Differences and Improvements



#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

#### 1. register all elements that need hyphenation

- 2. if the language is not set, ask for it

- execution is fast.
- downloading the script and the patterns takes time



#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

- 1. register all elements that need hyphenation
- 2. if the language is not set, ask for it
- 3. download the patterns, if not already done

- execution is fast.
- downloading the script and the patterns takes time

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

- 1. register all elements that need hyphenation
- 2. if the language is not set, ask for it
- 3. download the patterns, if not already done
- 4. split the paragraphs in words (and URLs)

- execution is fast.
- downloading the script and the patterns takes time

- 1. register all elements that need hyphenation
- 2. if the language is not set, ask for it
- 3. download the patterns, if not already done
- 4. split the paragraphs in words (and URLs)
- 5. process each word, put ­ at every valid breakpoint
- execution is fast.
- downloading the script and the patterns takes time

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/



- 1. register all elements that need hyphenation
- 2. if the language is not set, ask for it
- 3. download the patterns, if not already done
- 4. split the paragraphs in words (and URLs)
- 5. process each word, put ­ at every valid breakpoint
- 6. The browser will re-render the text automatically, taking in account the soft hyphens.
- execution is fast.
- downloading the script and the patterns takes time

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

- 1. register all elements that need hyphenation
- 2. if the language is not set, ask for it
- 3. download the patterns, if not already done
- 4. split the paragraphs in words (and URLs)
- 5. process each word, put ­ at every valid breakpoint
- 6. The browser will re-render the text automatically, taking in account the soft hyphens.
- execution is fast.
- downloading the script and the patterns takes time

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/



- 1. register all elements that need hyphenation
- 2. if the language is not set, ask for it
- 3. download the patterns, if not already done
- 4. split the paragraphs in words (and URLs)
- 5. process each word, put ­ at every valid breakpoint
- 6. The browser will re-render the text automatically, taking in account the soft hyphens.
- execution is fast.
- downloading the script and the patterns takes time script: 25 KB, en: 25 KB, pl: 37 KB, de:

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/



#### Main Differences

# Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

#### don't care about space in RAM, care about program size

- no Trie (retrieval tree)
  - no special data structures in Javascript
  - using a trie is faster in execution (10ms)
  - but: building the tree from the patterns takes time
  - but: for a tree extra code is used (uses bandwith)
  - but: transferring the hardcoded trie is no solution, either (overhead: 50%)
- using a hash table (Javascript: object) instead
- ▶ UTF-8 (Thanks to Arthur and Mojca)

#### Motivation

layout w/o hyphenation layout with hyphenation

he TEX syphenation

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

Creating the patterns (patgen)

sing the patterns

HTML and the

The Port to

Server side or Client sid

ow it works

Differences and Improvements



#### Main Differences

- don't care about space in RAM, care about program size
- no Trie (retrieval tree)
  - no special data structures in Javascript
  - using a trie is faster in execution (10ms)
  - but: building the tree from the patterns takes time
  - but: for a tree extra code is used (uses bandwith)
  - but: transferring the hardcoded trie is no solution, either (overhead: 50%)
- using a hash table (Javascript: object) instead
- ▶ UTF-8 (Thanks to Arthur and Mojca)

## Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

he TEX
phenation

The original TEX hyphenation algorithm

he current TEX yphenation algorith (983)

Creating the patterns (patgen)

sing the patterns

HTML and the soft hyphen

The Port to

Server side or Client side?

Differences and Improvements



#### Main Differences

- don't care about space in RAM, care about program size
- no Trie (retrieval tree)
  - no special data structures in Javascript
  - using a trie is faster in execution (10ms)
  - but: building the tree from the patterns takes time
  - but: for a tree extra code is used (uses bandwith)
  - but: transferring the hardcoded trie is no solution, either (overhead: 50%)
- using a hash table (Javascript: object) instead
- ▶ UTF-8 (Thanks to Arthur and Mojca)

## Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

he TEX
phenation

The original TEX hyphenation algorith

he current TEX yphenation algorith 1983)

Creating the patterns patgen)

sing the patterns yphenation)

HTML and the soft hyphen

The Port to

Server side or Client side?

Differences and Improvements

Back to the Eutu

## Packing the patterns (helper: compressor):

- size of the pattern file does matter
- ▶ no whitespace (> 12% saved!)

```
a1 a1 e1 e1 i1 o1 ó1 u1 y1 _a1 _b8 _c8 _ć8 _d8
```

```
2:'a1ą1e1ę1i1o1ó1u1y1',
3:'_a1_b8_c8_ć8_d8_e1 f8
```

- http-requests take time
- merge the script and the necessary patterns (usualy just one) in one file
- ▶ saves 2 requests per pattern-file

## Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

lavout w/o hyphenation

The T<sub>E</sub>X hyphenation

The original TEX hyphenation algorithm

he current TEX
yphenation algorithm

(patron)

(patgen) Using the patterns

TML and the

soft hyphen

#### The Port to Javascript

Server side or Client side?

How it works

Differences and
Improvements

#### Packing the patterns (helper: compressor):

- size of the pattern file does matter
- ▶ no whitespace (> 12% saved!)

```
a1 a1 e1 e1 i1 o1 ó1 u1 y1 _a1 _b8 _c8 _ć8 _d8
```

```
2:'a1a1e1e1i1o1ó1u1y1',
3:'_a1_b8_c8_ć8_d8_e1_f8
```

- http-requests take time
- merge the script and the necessary patterns (usualy just one) in one file
- saves 2 requests per pattern-file

## Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o nypnenation

The T<sub>E</sub>X hyphenation algorithm

The original TEX hyphenation algorithm

The current TEX hyphenation algorithm 1983)

Creating the patterns (patgen)

sing the patterns yphenation)

HTML and the soft hyphen

The Port to

Server side or Client side?

Differences and Improvements

#### Packing the patterns (helper: compressor):

- size of the pattern file does matter
- ▶ no whitespace (> 12% saved!)

```
a1 q1 e1 e1 i1 o1 o1 u1 y1 _a1 _b8 _c8 _c8 _d8
```

```
2: 'alalelelilolólulyl',
```

## Merging script and patterns in one file (helper: merge+pack)

- http-requests take time
- merge the script and the necessary patterns (usualy just one) in one file
- ▶ saves 2 requests per pattern-file

## Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

The T<sub>E</sub>X hyphenation algorithm

The original TEX hyphenation algorithm

he current TEX phenation algorithr 983)

Creating the patterns (patgen)

Using the patterns (hyphenation)

HTML and the

The Port to

Server side or Client side?

Differences and Improvements

lack to the Euture

#### Packing the patterns (helper: compressor):

- size of the pattern file does matter
- ▶ no whitespace (> 12% saved!)

```
a1 a1 e1 e1 i1 o1 ó1 u1 y1 _a1 _b8 _c8 _ć8 _d8
```

```
2: 'alalelelilolólulyl',
```

Merging script and patterns in one file (helper: merge+pack)

- http-requests take time
- merge the script and the necessary patterns (usualy just one) in one file
- saves 2 requests per pattern-file

## Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm

The current TEX hyphenation algorithn (1983)

(patgen)

Using the patterns (hyphenation)

## HTML and the soft hyphen

he Port to

Server side or Client side?

Differences and

Improvements

4 D > 4 B > 4 B > 4 B > B B 9 9 9

# Using reduced pattern sets for static sites (helper: reducePatternSet)

- most patterns are not used
- ▶ if the the text will not change, use a precomputed subset
- savings vary
- ▶ only take in account breakpoints of composite words: Zeilen-ende instead of Zei-len-en-de de patterns are now 37 KB instead of 74 KB 265683 good, 22837 bad, 995752 missed 21.06 %, 1.81 %, 78.94 %
- or use different settings for patgen!

## Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation

#### The T<sub>E</sub>X hyphenation

The original TEX hyphenation algorithm

(1977)
The current TEX
hyphenation algorithm

ypnenation algorithm 1983) Creating the patterns

(patgen)
Using the patterns

TML and the

HTML and the soft hyphen

#### The Port to Javascript

Server side or Client side?

#### Differences and Improvements

and the fire of

# Using reduced pattern sets for static sites (helper: reducePatternSet)

- most patterns are not used
- ▶ if the the text will not change, use a precomputed subset
- savings vary
- ▶ only take in account breakpoints of composite words: Zeilen-ende instead of Zei-len-en-de de patterns are now 37 KB instead of 74 KB 265683 good, 22837 bad, 995752 missed 21.06 %, 1.81 %, 78.94 %
- or use different settings for patgen!

## Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

he T<sub>E</sub>X yphenation lgorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

Creating the patterns (patgen)

Jsing the patterns hyphenation)

HTML and the soft hyphen

The Port to Javascript

Server side or Client side?

Differences and Improvements

improvements

#### Using reduced pattern sets for static sites (helper: reducePatternSet)

- most patterns are not used
- if the the text will not change, use a precomputed subset
- savings vary

#### (Recompute the patterns)

- only take in account breakpoints of composite words:
- or use different settings for patgen!

#### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Differences and

Improvements

Using reduced pattern sets for static sites (helper: reducePatternSet)

- most patterns are not used
- ▶ if the the text will not change, use a precomputed subset
- savings vary

(Recompute the patterns)

- only take in account breakpoints of composite words: Zeilen-ende instead of Zei-len-en-de de patterns are now 37 KB instead of 74 KB 265683 good, 22837 bad, 995752 missed 21.06 %, 1.81 %, 78.94 %
- or use different settings for patgen!

Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm

he current TEX hyphenation algorithm 1983)

1983)
Creating the patters
(natgen)

Jsing the patterns

ITML and the

The Port to

Server side or Client side?

Differences and

Improvements

# Using reduced pattern sets for static sites (helper: reducePatternSet)

- most patterns are not used
- ▶ if the the text will not change, use a precomputed subset
- savings vary

#### (Recompute the patterns)

- only take in account breakpoints of composite words: Zeilen-ende instead of Zei-len-en-de de patterns are now 37 KB instead of 74 KB 265683 good, 22837 bad, 995752 missed 21.06 %, 1.81 %, 78.94 %
- or use different settings for patgen!

## Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The T<sub>E</sub>X hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

1983) Creating the patterns (patgen)

Jsing the patterns

HTML and the

The Port to

avascript erver side or Client si

ow it works

Differences and Improvements



- Problems upon copy/paste of hyphenated text
- Problems with loaded fonts (@font-face
- ▶ Patterns very different in size
- some rare misplaced hyphenation breaks may happen

## Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

#### Motivation

layout w/o hyphenation layout with hyphenation

#### The TEX hyphenation algorithm

The original TEX hyphenation algorith

hyphenation algorithm (1977)

The current TEX hyphenation algorithm 1983)

Creating the patterns (patgen)

hyphenation)

HTML and the soft hyphen

## he Port to

erver side or Client sid

How it works
Differences and
Improvements

Back to the Futur

4 D > 4 A > 4 E > 4 E > E | E | 9 Q (> |

- Problems upon copy/paste of hyphenated text
- ▶ Problems with loaded fonts (@font-face)
- ► Patterns very different in size
- some rare misplaced hyphenation breaks may happen

## Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX

Creating the pattern (patgen)

Ising the patterns

HTML and the

he Port to

Server side or Client side? How it works

How it works
Differences and
Improvements



- Problems upon copy/paste of hyphenated text
- ▶ Problems with loaded fonts (@font-face)
- Patterns very different in size
- some rare misplaced hyphenation breaks may happen

## Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivatio

layout w/o hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

(1983) Creating the pattern

(patgen)

Jsing the patterns hyphenation)

HTML and the

he Port to

Server side or Client side

How it works

Differences and

Improvements



- Problems upon copy/paste of hyphenated text
- ▶ Problems with loaded fonts (@font-face)
- Patterns very different in size
- some rare misplaced hyphenation breaks may happen

### Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The T<sub>E</sub>X hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

1983) Creating the patterns

patgen) Ising the patterns

yphenation)

HTML and the soft hyphen

The Port to

erver side or Client side

Differences and

Improvements



- CSS3: browsers do hyphenation (w/o hyphenator.js)

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/



- CSS3: browsers do hyphenation (w/o hyphenator.js)
- ► TUG: maintained hyphenation patterns (beware of size!)

#### Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/



- CSS3: browsers do hyphenation (w/o hyphenator.js)
- ► TUG: maintained hyphenation patterns (beware of size!)
- Wish: better typography in web sites.
- ▶ Me: Try to rewrite PatGen for UTF-8 Support

## Hyphenation for HTML

Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation layout with hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

The current TEX hyphenation algorithm

1983) Creating the patterns

(patgen) Using the patterns

(hyphenation)

HTML and the

he Port to

Server side or Client side? How it works

Differences and



- CSS3: browsers do hyphenation (w/o hyphenator.js)
- ► TUG: maintained hyphenation patterns (beware of size!)
- Wish: better typography in web sites.
- ▶ Me: Try to rewrite PatGen for UTF-8 Support

## Hyphenation for HTML

#### Mathias Nater mathias@mnn.ch http://mnn.ch/

Motivation

layout w/o hyphenation

The TEX hyphenation algorithm

The original TEX hyphenation algorithm (1977)

> The current TEX hyphenation algorithm (1983)

Creating the pattern (patgen)

Using the patterns

HTML and the

soft hyphen

The Port to lavascript

Server side or Client side How it works

Differences and



In chapter 9.3.3. the HTML 4.01 Specification tells us the following about hyphenation in HTML:

[...] The soft hyphen tells the user agent where a line break can occur.

[...] If a line is broken at a soft hyphen, a hyphen character must be displayed at the end of the first line. If a line is not broken at a soft hyphen, the user agent must not display a hyphen character. [...] The soft hyphen is represented by the character entity reference ­ (­ or ­)

Return

 ${\sf Appendix}$ 

```
hyphenation
.hyphenation.
                2i o
          1n a
                   o2n
       h e2n
           n2a t
              1t i o
h e n a4
  h y3p h
       h e n5a t
. h y3p h e2n5a4t2i o2n .
hy-phen-ation
```

Appendix

ightharpoonup David Antoš (2001):

PatLib, Pattern Manipulating Library – Master Thesis Masaryk University Brno, Faculty of Informatics

Donald E. Knuth (1999):

Digital Typography.

Stanford, California: Center for the Study of Language and Information ISBN 1-57586-010-4

Franklin Mark Liang (1983):

Word Hy-phen-a-tion by Com-put-er. PhD thesis Department of Computer Science, Stanford University: Stanford, CA 94305.

http://www.tug.org/docs/liang/liang-thesis.pdf

Appendix

CHRISTINE RÖMER, HERBERT VOSS (2008):

Deutsche Silbentrennmuster – aus linguistischer und TEXnischer Sicht. PDF, Jena 06. 03. 2008 http://www.personal.uni-

jena.de/~xcr/v2/Dateien/File/Jena2008.pdf

RAGGETT DAVE, LE HORS ARNAUD, JACOBS IAN (1999):

HTML 4.01 Specification – W3C Recommendation 24 December 1999.

http://www.w3.org/TR/html401/