

The T_EX hyphenation applied to HTML

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

Mathias Nater
mathias@mn.ch
<http://mn.ch/>

BachoT_EX 2010

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Organisation

Motivation

Text layout without hyphenation

Text layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX hyphenation algorithm (1977)

The current T_EX 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Organisation

Motivation

Text layout without hyphenation

Text layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX hyphenation algorithm (1977)

The current T_EX 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Organisation

Motivation

Text layout without hyphenation

Text layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX hyphenation algorithm (1977)

The current T_EX 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Organisation

Motivation

Text layout without hyphenation

Text layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX hyphenation algorithm (1977)

The current T_EX 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

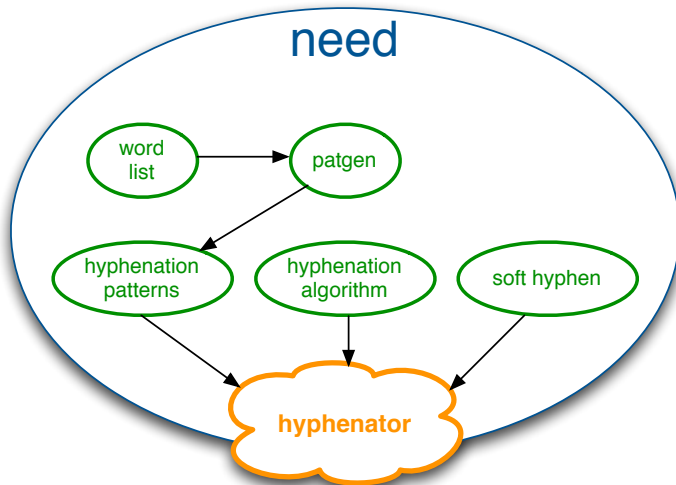
Differences and
Improvements

Back to the Future

Organisation

Hyphenation for HTML

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



Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX hyphenation algorithm (1977)

The current T_EX 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

Text layout without hyphenation

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Text layout without hyphenation

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Text layout without hyphenation

Current Browsers

- ▶ MS IE 6/7/8 (~ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Text layout without hyphenation

Current Browsers

- ▶ MS IE 6/7/8 (~ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Text layout without hyphenation

Current Browsers

- ▶ MS IE 6/7/8 (~ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Text layout without hyphenation

text-align: left;

Hyphenation for
HTML

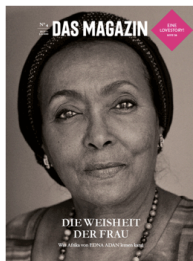
Mathias Nater
mathias@mnn.ch
<http://mnn.ch/>

DAS MAGAZIN

2010/4 Letztes Heft | Archiv | Nächstes Heft

Anmelden | Mein Profil

Suchen



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 Kopfhäger, 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

HIER FRAGT DER CHEF

Finn Canonica auf DRS 4 Talk

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 soft hyphen

The Port to Javascript

Server side or Client side?

How it works

Differences and Improvements

Back to the Future

Text layout without hyphenation

text-align: justify;

Hyphenation for
HTML

Mathias Nater
mathias@mn.ch
<http://mn.ch/>

DAS MAGAZIN

2010/4 Letztes Heft | Archiv | Nächstes Heft

Anmelden | Mein Profil

Suchen



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 Kopfgänger, 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

HIER FRAGT DER CHEF

Finn Canonica auf DRS 4 Talk

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 soft hyphen

The Port to Javascript

Server side or Client side?

How it works

Differences and Improvements

Back to the Future

Text layout with hyphenation

text-align: left;

Hyphenation for
HTML

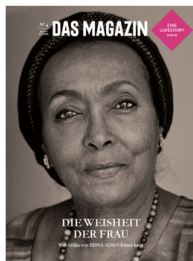
Mathias Nater
mathias@mnn.ch
<http://mnn.ch/>

DAS MAGAZIN

2010/4 Letztes Heft | Archiv | Nächstes Heft

Anmelden | Mein Profil

Suchen



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 Kopfhänger, 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

HIER FRAGT DER CHEF

Finn Canonica auf DRS 4 Talk

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 soft hyphen

The Port to Javascript

Server side or Client side?

How it works

Differences and Improvements

Back to the Future

Text layout with hyphenation

text-align: left;

Hyphenation for
HTML

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

DAS MAGAZIN

2010/4 Letztes Heft | Archiv | Nächstes Heft

Anmelden | Mein Profil

Suchen



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 Kopfküßler, 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

HIER FRAGT DER CHEF

Finn Canonica auf DRS 4 Talk

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 soft hyphen

The Port to Javascript

Server side or Client side?

How it works

Differences and Improvements

Back to the Future

Text layout with hyphenation

text-align: justify;

Hyphenation for
HTML

Mathias Nater
mathias@mn.ch
<http://mn.ch/>

DAS MAGAZIN

2010/4 Letztes Heft | Archiv | Nächstes Heft

Anmelden | Mein Profil

Suchen



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 Kopfpäger, 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

HIER FRAGT DER CHEF

Finn Canonica auf DRS 4 Talk

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 soft hyphen

The Port to Javascript

Server side or Client side?

How it works

Differences and Improvements

Back to the Future

Text layout with hyphenation

text-align: justify;

Hyphenation for
HTML

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

DAS MAGAZIN

2010/4 Letztes Heft | Archiv | Nächstes Heft

Anmelden | Mein Profil

Suchen



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 Kopffäger, 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

HIER FRAGT DER CHEF

Finn Canonica auf DRS 4 Talk

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 soft hyphen

The Port to Javascript

Server side or Client side?

How it works

Differences and
Improvements

Back to the Future

We *need* (automatic) hyphenation in HTML!

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The original T_EX 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The original T_EX 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The original T_EX 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The original T_EX 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The original T_EX 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The original T_EX 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The original T_EX 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The current T_EX 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 (T_EX)
- ▶ 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

**The current T_EX
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

The current T_EX 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 (T_EX)
- ▶ 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The current T_EX 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 (T_EX)
- ▶ 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

The current T_EX 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 (T_EX)
- ▶ 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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Creating patterns with patgen

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Creating patterns with patgen

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Creating patterns with patgen

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Creating patterns with patgen

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Creating patterns with patgen

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Creating patterns with patgen

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns

- ▶ .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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns

- ▶ .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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns

- ▶ .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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns

- ▶ .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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns

- ▶ .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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns (example)

incredible

```
. i n c r e d i b l e .  
. i n1  
b2l2  
4e d i  
i1b l  
n1c r  
b4l e .  
5c r e d  
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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns (example)

incredible

. i n c r e d i b l e .

. i n1

b2l2

4e d i

i1b l

n1c r

b4l e .

5c r e d

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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns (example)

incredible

. i n c r e d i b l e .

. i n1

b2l2

4e d i

i1b l

n1c r

b4l e .

5c r e d

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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns (example)

incredible

. i n c r e d i b l e .

. i n1

b2l2

4e d i

i1b l

n1c r

b4l e .

5c r e d

e d3i b

2r2e d

. i n5c2r4e d3i1b4l2e .

in-cred-i-ble

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

applying the patterns (example)

incredible

. i n c r e d i b l e .

. i n l

b2l2

4e d i

i1b l

n1c r

b4l e .

5c r e d

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

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

► example 2

HTML and the Soft Hyphen

- ▶ limited control over textlayout
(text-align: left | right | justify)
- ▶ manual line breaks (
)
- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

HTML and the Soft Hyphen

- ▶ limited control over textlayout
(`text-align: left | right | justify`)
- ▶ manual line breaks (`
`)
- ▶ 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

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 soft hyphen

The Port to Javascript

Server side or Client side?

How it works

Differences and
Improvements

Back to the Future

HTML and the Soft Hyphen

- ▶ limited control over textlayout
(`text-align: left | right | justify`)
- ▶ manual line breaks (`
`)
- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

HTML and the Soft Hyphen

- ▶ limited control over textlayout
(`text-align: left | right | justify`)
 - ▶ manual line breaks (`
`)
 - ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

HTML and the Soft Hyphen

- ▶ limited control over textlayout
(text-align: left | right | justify)
- ▶ manual line breaks (
)
- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Text layout with hyphenation

text-align: justify;

Hyphenation for
HTML

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

DAS MAGAZIN

2010/4 Letztes Heft | Archiv | Nächstes Heft

Anmelden | Mein Profil

Suchen



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 Kopfgänger, 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

HIER FRAGT DER CHEF

Finn Canonica auf DRS 4 Talk

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 soft hyphen

The Port to Javascript

Server side or Client side?

How it works

Differences and Improvements

Back to the Future

Text layout with hyphenation

text-align: justify;

Hyphenation for
HTML

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

DAS MAGAZIN

2010/4 Letztes Heft | Archiv | Nächstes Heft

Anmelden | Mein Profil

Suchen



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 Kopffäger, 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

HIER FRAGT DER CHEF

Finn Canonica auf DRS 4 Talk

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 soft hyphen

The Port to Javascript

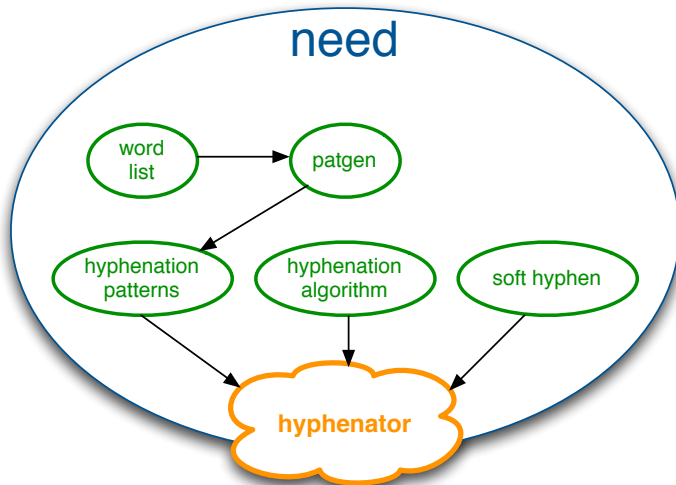
Server side or Client side?

How it works

Differences and Improvements

Back to the Future

Putting all together



Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Server side or Client side hyphenation?

Hyphenation for
HTML

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

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Server side or Client side hyphenation?

Hyphenation for
HTML

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

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Server side or Client side hyphenation?

Hyphenation for
HTML

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

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Server side or Client side hyphenation?

Hyphenation for
HTML

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

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Server side or Client side hyphenation?

Hyphenation for
HTML

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

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Server side or Client side hyphenation?

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Server side or Client side hyphenation?

Hyphenation for
HTML

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

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Server side or Client side hyphenation?

Hyphenation for
HTML

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

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Server side or Client side hyphenation?

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

My Decision

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

My Decision

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

My Decision

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

My Decision

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

My Decision

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

My Decision

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

My Decision

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

My Decision

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

How it works

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: 74 KB

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

How it works

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: 74 KB

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

How it works

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: 74 KB

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

How it works

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: 74 KB

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

How it works

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: 74 KB

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

How it works

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: 74 KB

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

How it works

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: 74 KB

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

How it works

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: 74 KB

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

How it works

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: 74 KB

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 soft hyphen

The Port to Javascript

Server side or Client side?

How it works

Differences and
Improvements

Back to the Future

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 bandwidth)
 - ▶ 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

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

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 bandwidth)
 - ▶ 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

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

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 bandwidth)
 - ▶ 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

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Improvements I

Packing the patterns (helper: compressor):

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

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

```
2: 'a1a1e1e1i1o1o1u1y1',
```

```
3: '_a1_b8_c8_c8_d8_e1_f8
```

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Improvements I

Packing the patterns (helper: compressor):

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

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

```
2: 'a1a1e1e1i1i1o1o1u1y1',
```

```
3: '_a1_b8_c8_c8_d8_e1_f8
```

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Improvements I

Packing the patterns (helper: compressor):

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

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

```
2: 'a1a1e1e1i1o1o1u1y1',
```

```
3: '_a1_b8_c8_c8_d8_e1_f8
```

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

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Improvements I

Packing the patterns (helper: compressor):

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

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

```
2: 'a1a1e1e1i1i1o1o1u1y1',
```

```
3: '_a1_b8_c8_c8_d8_e1_f8
```

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

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Improvements II

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!

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Improvements II

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!

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Improvements II

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

- ▶ most patterns are not used
- ▶ if 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!

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Improvements II

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!

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Improvements II

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!

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Hyphenator.js Problems and Oddities

Hyphenation for
HTML

Mathias Nater
mathias@mn.ch
<http://mn.ch/>

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Hyphenator.js Problems and Oddities

Hyphenation for
HTML

Mathias Nater
mathias@mn.ch
<http://mn.ch/>

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Hyphenator.js Problems and Oddities

Hyphenation for
HTML

Mathias Nater
mathias@mn.ch
<http://mn.ch/>

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

Hyphenator.js Problems and Oddities

Hyphenation for
HTML

Mathias Nater
mathias@mn.ch
<http://mn.ch/>

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

What the future shall/may bring

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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

What the future shall/may bring

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

What the future shall/may bring

Hyphenation for
HTML

Mathias Nater
mathias@mn.ch
<http://mn.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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

What the future shall/may bring

- ▶ 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

Motivation

layout w/o hyphenation

layout with hyphenation

The T_EX hyphenation algorithm

The original T_EX
hyphenation algorithm
(1977)

The current T_EX
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

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 `­`);

applying the patterns – example 2

Appendix

hyphenation

. h y p h e n a t i o n .

2i o

1n a

 o_{2n}
$$h_{e2n}$$

n2a t

1t i o

h e n a4

h y3p h

h e n5a t

. h y3p h e2n5a4t2i o2n .

hy-phen-ation

For Further Reading I

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



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



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



Department of Computer Science, Stanford University:
Stanford, CA 94305.
<http://www.tug.org/docs/liang/liang-thesis.pdf>

For Further Reading II



CHRISTINE RÖMER, HERBERT VOSS (2008):
Deutsche Silbentrennmuster – aus linguistischer und
T_EXnischer 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/>