人間の悩みを軽くする XML その過去と将来

佐々木　フェリクス,
<fsasaki@w3.org>

英語のスライド

日本語のスライド

データーの一例

「若きウェルテルの悩み」から（ドイツ語の原作、日本語訳、英語訳）

> Bester Freund, was ist das Herz des Menschen!

> Dear friend, what a piece of work is the human heart!

> 愛するともよ、人間の心なんて、なんと言う妙なものだろう！

修士論文の課題：「若きウェルテルの悩み」の多言語のコーパス、その作成、分析、表示など

若きウェルテルの悩みを軽くしてくれた XML 1.0 の性質

宣言型、分かりやすい構造 (木構造) 、実装の簡単な XML

```
<corpus>
 <s xml:lang="de">Bester Freund, was ist das Herz des Menschen!</s>
 <s xml:lang="en">Dear friend, what a piece of work is the human heart!</s>
 <s xml:lang="ja">愛するともよ、人間の心なんて、なんと言う妙なものだろう！</s>
</corpus>
```

若きウェルテルの悩みを軽くしてくれた XML 1.0 の性質

- 妥当性の検討

> 言語コーパスが多言語の文章の集合であることは XML 文書の妥当性に基づき、検討できる

- XPath、XSLT、XSL-FO、XQueryなどに基づき、コーパスの問い合わせ、構造の変換、表示が可能になる

```
for $japaneseTranslations in document("corpus.xml")//s
where @xml:lang='ja'
return $japaneseTranslations
```

- 若きウェルテルのコーパスを Web に乗せ、XMLによる Webサービスでアクセスできる

私の悩みを軽くしてくれた XML

- コーパスデータの妥当性を簡単に検討することができるようになった！
- 単語を数えるなどの手間作業が自動化で楽になった
- 一般的な組版が簡単になった

若きウェルテルの悩みをもっと軽くしてくれる、これからの XML

- XML 1.0 5th edition は XML names の問題になったユニコードバージョンへの依存性がなくなる

```
<corpus
author="ヨハン・ヴォルフガング・フォン・ゲーテ"> ...</corpus>
```

(author属性は IDREF型)

- XML Schema 1.1 は新たな妥当性の機能を導入し、文脈により、要素などの型を当て嵌められる。(例えば日本語の文書、英語の文書、又はドイツ語の文書の相違点をスキーマでも定義できるようになる)

```
<corpus>
 <s xml:lang="de">Bester Freund, was ist das Herz des Menschen!</s>
 <s xml:lang="en">Dear friend, what a piece of work is the human heart!</s>
 <s xml:lang="ja">愛するともよ、人間の心なんて、なんと言う<ruby>...</ruby>ものだろう！</s>
</corpus>
```

若きウェルテルの悩みをもっと軽くしてくれる、これからの XML

- XSL-FO 2.0 **における、日本語組版の新たな機能**

> 「若きウェルテルの悩み」という小説を縦組みにし、 Web でも日本人にとって読みやすくなる

- XQuery 1.0 and XPath 2.0 Full-Text 1.0

> 言語ごとに文書解析（異形態から語幹の抽出、 stop words の除外、類語辞典の使用など）

- XProc: An XML Pipeline Language

> コーパスの妥当性の検討 > コーパスの分析 > 分析の結果を表示（組版）

XML は、若きウェルテルの悩みを完全になくすことは…

…できなかったが、たくさんの人を助けることができた

. (コーパス) データーの統合性、又データーの持続可能性が改良された

. (コーパス) データーの処理を楽にしてくれた大勢の XML ツール

. (コーパス) データーの交換の可能性が改良された

XML 10 @ W3C を一緒に祝いしましょう

- http://www.w3.org/2008/xml10
- **グリーティングカードを送りましょう**
  http://www.w3.org/2008/xml10/card/greeting-form
- XML 10 **のロゴをブログなどでも使いましょう**