

التكنولوجيا وراء اللغة العربية

いなにわうどん (@kyoto_inaniwa)

2024/5/11 UNTIL.LT #0x04

التكنولوجيا وراء اللغة العربية

いなにわうどん (@kyoto_inaniwa)

2024/5/11 UNTIL.LT #0x04



いなにわうどん @kyoto_inaniwa

<https://いなにわうどん.みんな>

興味 組版, Web フロントエンド, HCI

キロロ

2020-2022

© いなにわうどん. ページ中の写真は クリエイティブ・コモンズ 表示 4.0 国際 (CC BY 4.0) の下で頒布しています。

2020 年度 未踏 IT 人材発掘・育成事業

和田 優斗 (横浜市立横浜サイエンスフロンティア)

強力なグラフィック機能を備えた 日本語組版処理システムの開発

— Twilight: An implementation of Graphical Typesetting —

Twilight は、XML、CSS を組版用に改良した言語を用いて、
テキストベースでグラフィカルな組版を実現する組版処理システムです。

▼実際の



1A- グリッドレイアウトに基づく組版作業を支援する 06 手書きストロークを入力としたインタフェース

和田 優斗, 志築文太郎 (筑波大学)

1. 研究背景・目的

グリッドレイアウト (図 1)

- 紙面を水平・垂直方向に分割した直線上にテキスト・図版を配置するデザイン手法
- グリッドレイアウトの活用により、デザイン 4 原則の「整列」「反復」を満たしたレイアウトを構築できる
- 既存のインタフェース (バウンディングボックス等)
- 整列等、難易度の高い操作が要求される

本研究の目的
グリッドデザインに則った文書のレイアウト作業を
簡便に実現するインタフェースの開発

2. 提案インタフェース

手書きによるストローク入力 (図 2) を用いて、グリッドレイアウトの生成 (図 3) を行うインタフェースを開発

提案インタフェースの特徴

- ストロークで構成されるラフスケッチからグリッドレイアウトを生成
- ラフスケッチはペンの種類やストローク形状 (図 4) で区別される
- レイアウト、テキスト、スタイルを分離して管理
- ドラッグ操作によって文字スタイルを WYSIWYG に調節可能 (図 5)
- スタイル (デフォルト・特設・有線等) はグリッドに別途割り当て (図 6)

図 4 入力として与えられるストロークの種類と機能

● テキストの編集

図 5 入力として与えられるストロークの種類と機能



Twilight

2020 年度未踏事業

組版の実装は
たのしい!!

gridstroke

インタラクシオン 2023

雙峰祭Web サイト

2022

漫画用の組版システム

業務委託 (2021-)

組版のたのしさ（むずかしさ）

世界中の様々な

言語 + 組版規則 +

フォント + フォーマット

ところが……

世界には 左→右（LTR）の言語 しか考慮しない人が多い！！

日本語組版とは？

あの日から
ずっとずっと
後悔ばかり
押しよせて

混植

アンチゴチ（明朝＋ゴシック）

み　とう　じ　ぎょう

未踏事業（2000年に開始された情報処理推進機構による人材発掘育成事業）

モノルビ

割注

時に、初春の令月にして、気淑く和らぎ
時に、初春の令月にして、気淑く和らぎ

カーニング

私は「あいう。」

私はhumanです

「醒めた面、（つ

文字アキ量設定

要池木小

人名字取り

上村拓一
泉純一

潤彰哉郎

から引用された。

日から現在に至る。「万葉集」

の元号。2019年5月1

平成の後。大化以降

248
縦中横

令和は、日本の元号の一つ。

葛飾区
葛飾区

異体字

追い込み

ぶらさがり

An aerial photograph of Dubai, United Arab Emirates, featuring the Burj Khalifa as the central landmark. The city's dense urban landscape is visible, surrounded by desert terrain and some water bodies in the distance.

アラビア語

世界第3位

約2.4億人



幣束

@goshuinchou

フォローする



パリーグすげえ試合してんな

2023年10月18日 20:32 · 3,444 表示

10 いいね



🔵 確証あり | مُروا المتيته @ar_sh9 · 16分



返信先: @goshuinchouさん

それが分かりません、どういう意味ですか？



🔵 確証あり | مُروا المتيته @ar_... · 8時間

إلي يبغي يرفع مشاهداته و يشترك في أرباح



خذوا هذا الحساب الياباني

(@TOKAI_ONAIR) أقسم بالله جبت منه

أكثر من 1.5 مليون مشاهد بتفاعل ساعة و...



وَاللّٰهُمَّ صَلِّ وَسَلِّمْ

アッラーが彼とその一族のうえに
祈りと平安を与えますように

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

慈悲あまねきアッラーの御名において

アラビア語のむずかしさ

読むために最低限すべきことが多い

① 双方向処理

② 字形処理

① 双方向処理

漢文のように返って読む

- ・ 欧文／和文／数字：LTR、アラビア語：RTL



Unicode BiDi Algorithm

BiDi = BiDirectional の略

- グリフごとに書字方向が定義される
- 連続するグリフ毎でグループ化
→ 埋め込みレベルを表現
- 埋め込みレベル毎に LTR, RTL を決定

参考

- UAX #9: Unicode Bidirectional Algorithm
<https://unicode.org/reports/tr9/>
- 右から左に書く言語を支える技術
<https://note.com/ttuusskk/n/ne1f4466bb45f>

Unicode Standard Annex #9

THE BIDIRECTIONAL ALGORITHM

Version	Unicode 4.0.0
Authors	Mark Davis (mark.davis@us.ibm.com)
Date	2003-04-17
This Version	http://www.unicode.org/reports/tr9/tr9-11.html
Previous Version	http://www.unicode.org/reports/tr9/tr9-10.html
Latest Version	http://www.unicode.org/reports/tr9/
Tracking Number	11

Summary

This document describes specifications for the positioning of characters flowing from right to left, such as Arabic or Hebrew.

Status

*This document has been reviewed by Unicode members and other interested parties, and has been approved by the Unicode Technical Committee as a **Unicode Standard Annex**. This is a stable document and may be used as reference material or cited as a normative reference by other specifications.*

A Unicode Standard Annex (UAX) forms an integral part of the Unicode Standard, but is published as a separate document. The Unicode Standard may require conformance to normative content in a Unicode Standard Annex, if so specified in the Conformance chapter of that version of the Unicode Standard. The version number of a UAX document corresponds to the version number of the Unicode Standard

② 字形処理

文字がどこに現れるかで変わる

- isolated, initial, medial, final と 4 つの形状
- 隣接（接続）する文字によって変化

合字

- 隣接するグリフが結合し、
新たなグリフに置き換わる
- Unicode レベルで処理
- 欧文 (fi, ff) にも共通する概念

No	Name	Isolate	Beginning	Middle	End
1	Alif	ا	-	-	آ
2	Baa	ب	ب	ـ	ـب
3	Taa	ت	ت	ـ	ـت
4	Thaa	ث	ث	ـ	ـث
5	Jeem	ج	ج	ـ	ـج
6	Haa	ح	ح	ـ	ـح
7	Khaa	خ	خ	ـ	ـخ
8	Daal	د	-	-	ـ
9	Dhal	ذ	-	-	ـذ
10	Raa	ر	-	-	ـر
11	Zaa	ز	-	-	ـز
12	Seen	س	س	ـ	ـس
13	Sheen	ش	ش	ـ	ـش

出典：Ahmed Lawgali,
A Survey on Arabic Character Recognition.

Correct

لا ← ا + ج
U+0627 U+0644

Wrong

U ← l + J ← a + J
Final shape Initial shape U+0627 U+0644

出典：<https://www.w3.org/TR/alreq/>

むすびに

アラビア語は続くよどこまでも

- ・ミラーリング、カシーダ、結合文字、改行……

組版たのしい！！

- ・身近な日本語でも知らない組版規則は山程ある
- ・葛藤もある

「言語を知らずして組版の実装は可能なのか……？」