ترتیب‌بندی و مرتب‌سازی برای زبان فارسی ایران، پیشنویس ۵

Copyright © 2005 Sharif FarsiWeb, Inc.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

کلیهٔ حقوق برای شرکت فارسی‌وب شریف (سهامی خاص) محفوظ است.

اجازهٔ تکثیر، توزیع، و/یا تغییر این مستند تحت شرایط اجازه‌نامهٔ مستندات آزاد گنو (نسخهٔ 1.2 یا هر نسخهٔ جدیدتری که توسط بنیاد نرم‌افزارهای آزاد منتشر شود) داده می‌شود؛ بدون «قسمتهای بی‌تغییر»، «متن روی جلد»، یا «متن پشت جلد». یک نسخه از اجازه‌نامهٔ مذکور در بخش «اجازه‌نامهٔ مستندات آزاد گنو» آمده است.

۱ مقدمه

مرتب‌سازی، نیاز اولیهٔ تهیهٔ هرگونه فهرست کارآمد است. ولی مرتب‌سازی رشته‌ها و عبارات، به شکلی که هم خوانندهٔ آشنا به زبان بتواند به آسانی عبارت مورد نظرش را بیابد، و هم مطابق توصیه‌های لغت‌نامه‌نویسان، زبان‌شناسان، کتابداران، و نمایه‌سازان باشد، کاری نابدیهی است که به پردازش زبان طبیعی نیاز دارد. مثلاً در بعضی قوانین مرتب‌سازی، اگر عنوان کتابی «۲۰۰۱، ادیسهٔ فضایی» باشد، طوری مرتب می‌شود که انگار عنوانش «دو هزار و یک، ادیسهٔ فضایی» است؛ یا در بعضی قوانین نمایه‌سازی اگر عنوان مدخلی در نمایه «عدد π» باشد، طوری مرتب می‌شود که انگار عنوانش «عدد پی» است؛ یا در بعضی لغت‌نامه‌ها اگر دو کلمه از نظر حروف یکسان بوده و فقط از نظر اِعراب تفاوت داشته باشند، مرتب‌سازی بر اساس اِعراب صورت نمی‌گیرد بلکه بر اساس اهمیت کلمه برای خواننده صورت می‌گیرد.

این توصیف به قوانین ساده‌شدهٔ ترتیب‌بندی، بر اساس استاندارد بین‌المللی ISO/IEC 14651 و استاندارد فنی شمارهٔ ۱۰ یونی‌کد می‌پردازد. در این قوانین، فقط رشتهٔ ورودی در نظر گرفته می‌شود، نه اطلاعات جانبی دیگر و قوانین پیچیدهٔ ترتیب‌بندیهای خاص‌منظوره.

ترتیب‌بندی توصیف شده در این توصیف، تنها رشتهٔ یونی‌کدی متناظر کلید را در نظر می‌گیرد و اطلاعات معنایی احتمالی دیگر کلیدها در مرتب‌سازی آنها نقشی ندارند. مثلاً در توصیف حاضر، از آنجا که نویسهٔ اول رشتهٔ «۱۱»، رقم یک، قبل از نویسهٔ اول رشتهٔ «۲»، رقم ۲، است، رشتهٔ «۱۱» قبل از رشتهٔ «۲» قرار می‌گیرد. (بدیهی است می‌توان برای کاربردهایی که چیزی غیر از این رفتار در آنها مطلوب باشد، این الگوریتم را گسترش داد.)

این توصیف باید در مواردی که قانون دیگری برای ترتیب‌بندی فارسی مشخص نشده است استفاده شود. در کاربردهای خاصتر، بهتر است الگوریتم پیاده‌سازی‌شده، گسترشی از الگوریتم مشخص‌شده در این توصیف باشد. کاربردهایی که از قسمتهای اجباری الگوریتم فعلی پشتیبانی نکنند با این توصیف سازگار نیستند.

در این توصیف روش ترتیب‌بندی برای مرتب‌سازی متنهای فارسی بیان شده است، نه الگوریتمهای قابل استفاده برای مرتب‌سازی (که مربوط به دروس پایه‌ای علوم رایانه است).

۲ اصطلاحات و تعاریف

مرتب‌سازی، کلید، و رکورد

مرتب‌سازی (sorting) عملی است که فهرستی از چیزها را (که رکورد نامیده می‌شوند)، به ترتیب خاصی (مثلاً صعودی)، بر اساس بخش مشخصی از آن چیزها (که کلید نامیده می‌شوند)، قرار می‌دهد. مثلاً در مرتب‌سازی پرونده‌های پزشکی یک مطب، پرونده‌ها رکورد هستند و شمارهٔ آنها کلیدشان است.

در واقع هیچ اطلاعاتی از رکورد، غیر از کلید، در مرتب‌سازی استفاده نمی‌شود.

مرتب‌سازی ممکن است با الگوریتمهای مختلفی انجام شود که رشتهٔ مشخصی در علوم رایانه (computer science) است. از این الگوریتمها، Quick Sort معروفترین و معمولترین آنها است.

توصیف حاضر به الگوریتمهای مختلف قابل استفاده برای مرتب‌سازی، که مربوط به دروس پایهٔ علوم رایانه است، نمی‌پردازد. برای اطلاع دربارهٔ الگوریتمهای مختلف مرتب‌سازی، به جلد سوم کتاب دانلد کنوت، «*The Art of Computer Programming*» با عنوان فرعی «*Sorting and Searching*» مراجعه شود.

مرتب‌سازی مقایسه‌ای

مرتب‌سازی مقایسه‌ای به هر روش یا الگوریتمی برای مرتب‌سازی گفته می‌شود که فرض خاصی دربارهٔ محتویات کلیدها نکند و هر گونه تفسیر محتویات کلیدها در آن محدود به تعیین این مسئله باشد که بین دو کلید داده‌شده، کدام‌یک باید در یک فهرست مرتب‌شده قبل از دیگری بیاید. اکثر روشها و الگوریتمهای مرتب‌سازی از نوع مرتب‌سازی مقایسه‌ای هستند.

نویسه

نویسه معادل character در استاندارد یونی‌کد است.

دنباله، دنبالهٔ عددی

در این توصیف منظور از دنباله، فهرست متناهی و نمایه‌داری از عددها یا نویسه‌ها است. دنبالهٔ عددی دنباله‌ای از عددها است.

رشته

در این توصیف، به هر دنباله‌ای از نویسه‌های یونی‌کدی، رشته گفته می‌شود. در این توصیف فرض می‌شود که رشته‌ها از ترتیب معنایی (در چارچوب الگوریتم دوجهتهٔ یونی‌کد) پیروی می‌کنند.

ترتیب‌بندی

ترتیب‌بندی (collation) عملی است که ورودی آن دو دنباله است و خروجی آن حاصل مقایسهٔ آن دو دنباله برای تعیین این که کدام رشته باید در یک فهرست مرتب‌شده قبل از دیگری بیاید (حاصل مقایسه ممکن است «هم‌ارز» نیز باشد، که به این معنی است که ترتیب دو دنباله اهمیتی ندارد). ترتیب‌بندی در واقع همان عمل پایهٔ مورد استفاده برای مرتب‌سازیهای مقایسه‌ای است، در صورتی که کلیدها رشته یا دنبالهٔ عددی باشند.

قبل، بعد، هم‌ارز

از نظر ترتیب‌بندی دو دنباله «هم‌ارز» گفته می‌شوند اگر اهمیت و/یا تفاوتی نداشته باشد که کدام‌یک در یک فهرست مرتب‌شده قبل از دیگری بیاید.

اگر دو دنبالهٔ عددی هم‌ارز باشند حتماً برابر هم هستند (یعنی طول دو دنباله و تک‌تک اعضایشان برابرند) ولی این مسئله در مورد رشته‌ها برقرار نیست.

دنبالهٔ الف «قبل» از دنبالهٔ ب است، اگر لازم باشد که دنبالهٔ الف در یک فهرست مرتب‌شده قبل از دنبالهٔ ب بیاید. در این صورت گفته می‌شود دنبالهٔ ب «بعد» از دنبالهٔ الف است.

وزن

در ترتیب‌بندی، وزن عددی است که در هر یک از سطوح ترتیب‌بندی به نویسه‌ها تخصیص داده می‌شود تا بر اساس آن مقایسه شوند. مثلاً در سطح یک ترتیب بندی (حروف اصلی)، از آنجا که باید حرف «ب» بعد از حرف «الف» ترتیب‌بندی شود، به آن وزن بیشتری تخصیص یافته است، ولی از آنجا که تفاوت این دو حرف در سطح دوی ترتیب‌بندی اهمیتی ندارد، در آن سطح این دو نویسه وزن برابر دارند. برای اطلاع بیشتر از رفتار وزن، به الگوریتمهای این توصیف مراجعه کنید.

۳ توصیه‌ها و نیازهای پیاده‌سازی

لزومی ندارد که پیاده‌سازیهای این توصیف دقیقاً همان مراحل مشخص‌شده در این توصیف را انجام دهند. برای سازگاری با این توصیف کافی است خروجی ترتیب‌بندی هر دو رشتهٔ فارسی همواره و در همهٔ حالتها همان خروجی‌ای باشد که از این الگوریتم حاصل می‌شود.

این توصیف در مورد ترتیب مرتب‌سازی خطهای دیگر صحبت نمی‌کند و تنها به مرتب‌سازی متنهایی که به خط فارسی نوشته شده‌اند می‌پردازد. بنابراین در این توصیف فرض می‌شود که رشته‌های ورودی محدود به نویسه‌هایی هستند که در بخش «نویسه‌های پشتیبانی‌شدهٔ» این توصیف آمده‌اند. (باید توجه شود که اگر نویسه‌ای در آن بخش ذکر شده باشد ولی در هیچ‌یک از جدولهای ۱، ۲، و ۳ توصیف حاضر نیامده باشد، آن نویسه از نظر این توصیف در ترتیب‌بندی فارسی تقریباً قابل چشم‌پوشی است و وجود آن در رشته‌ها تقریباً تأثیری در ترتیب‌بندی آنها ندارد. البته این نویسه‌ها قابل حذف نیستند و وجود آنها در سطح سهٔ الگوریتم ترتیب‌بندی رشته‌ها مؤثر است.)

در صورتی که رشته‌ها نویسهٔ دیگری خارج از آن مجموعه داشته باشند، باید بر اساس استاندارد ISO/IEC 14651 یا استاندارد فنی شمارهٔ ۱۰ یونی‌کد، ولی با سازگاری کامل با ویژگیهای مشخص‌شده در این توصیف برای نویسه‌های پشتیبانی‌شده، ترتیب‌بندی شوند. سازگاری کامل به این معنی است که ترتیب‌بندی دو رشتهٔ فارسی، باید همواره همان ترتیب‌بندی حاصل از اجرای الگوریتمهای توصیف حاضر باشد (بدیهی است که وزنهای مشخص‌شده در این توصیف باید برای کار کردن با استانداردهای فوق، تبدیل شوند). برای دیدن مثالی از چنین پیاده‌سازیهایی، به پیاده‌سازی ترتیب‌بندی فارسی ایران در کتابخانهٔ C گنو (پروندهٔ fa_IR) مراجعه شود.

مؤکداً توصیه می‌شود که در صورت نیاز به پیاده‌سازی این توصیف، برای اطلاع دقیقتر از نیازهای ترتیب‌بندیهای استاندارد به استاندارد فنی شمارهٔ ۱۰ یونی‌کد و استاندارد بین‌المللی ISO/IEC 14651 مراجعه شود. مشخصاً، استاندارد فنی شمارهٔ ۱۰ یونی‌کد، شامل توصیه‌هایی برای کاهش زمان مقایسه برای ترتیب‌بندی است.

به‌علاوه، توصیه می‌شود که به خاطر احتمال پیش آمدن اشکالات معمول در پیاده‌سازی چنین الگوریتمهایی، که هر گونه اِشکال جزئی در پیاده‌سازی آنها باعث ناسازگاری پیاده‌سازی با این توصیف می‌شود، تا حد ممکن از پیاده‌سازیهای موجود الگوریتمهای ترتیب‌بندی استفاده شود. به‌عنوان مثال، می‌توان به کتابخانهٔ C گنو و کتابخانهٔ ICU شرکت آی‌بی‌ام اشاره کرد که هر دو نرم‌افزار آزاد هستند و مجوزشان نیز با GPL سازگار است. بدیهی است که لازم است جداول وزنها به شکل مناسب برای آن پیاده‌سازی، تبدیل شوند.

در هر صورت، مؤکداً پیشنهاد می‌شود که پیاده‌سازیها برای آزمایش پیاده‌سازیشان، اعم از پیاده‌سازی کامل یا تبدیل جداول، از داده‌های آزمون ضمیمهٔ این توصیف استفاده کنند. در این داده‌های آزمون، فهرست مرتبی از رشته‌ها که بر اساس توصیف حاضر ترتیب‌بندی شده‌اند، داده

شده است. پیاده‌سازیها باید بتوانند هر جایگشتی از آن فهرست را به همان ترتیب مشخص‌شده در آن فهرست مرتب کنند. هر پیاده‌سازی‌ای که جایگشتی از داده‌های آزمون ضمیمهٔ توصیف حاضر را به همان ترتیبی که در فهرست اولیه آمده‌اند مرتب نکند، با این توصیف سازگار نیست.

۴ شیوهٔ استفاده

کاربرهایی که بخواهند از ترتیب‌بندی مشخص شده در این الگوریتم برای مرتب‌سازی مقایسه‌ای رکوردهایی که کلیدشان یک رشتهٔ فارسی است استفاده کنند، کافی است الگوریتم مقایسهٔ کلیدشان را با الگوریتم مشخص‌شده در این توصیف جایگزین کنند. در اکثر زبانهای برنامه‌سازی، تابعی برای مرتب‌سازی، مثلاً بر اساس الگوریتم Quick Sort، موجود است که به‌عنوان آرگومان ورودی، یک تابع مقایسه می‌گیرد. در چنین مواردی می‌توان یک تابع ترتیب‌بندی سازگار با توصیف حاضر را به‌عنوان ورودی به چنین تابعی داد تا فهرستها را بر اساس آن تابع ترتیب‌بندی مرتب کند. برای مثال، در زبان برنامه‌سازی C تابعی به نام qsort وجود دارد که چنین کاری را می‌کند.

۵ نمادها

در این توصیف، نمادها به شکل تعیین شده در زیر به کار می‌روند:

منظور از U+20AC نویسهٔ «علامت یورو» از استاندارد یونی‌کد است که کد شانزده‌شانزدهی آن 20AC است.

منظور از U+06F0..U+06F9، همهٔ نویسه‌هایی از استاندارد یونی‌کد است که کُدشان از 06F0 تا 06F9 است، که شامل ده نویسه می‌شود (این نویسه‌ها ارقام فارسی هستند).

منظور از [۱، ۳، ۱۹، ۵] دنبالهٔ عددی‌ای از اعداد یک، سه، نوزده، و پنج است. در این توصیف، بر خلاف سنت معمول در متون ریاضیات و علوم رایانه، دنباله‌ها از راست به چپ نوشته می‌شوند و با ویرگول فارسی جدا می‌شوند.

منظور از <م، ـِ ل، ک> رشته‌ای از نویسه‌های «حرف فارسی میم»، «زیر فارسی»، «حرف فارسی لام»، و «حرف فارسی کاف» است. این رشته ممکن است برای خلاصه‌نویسی به شکل «مِلک» نیز نوشته شود. به همین ترتیب، منظور از <ن، ا، م، ه، فاصلهٔ مجازی، ه، ا> رشته‌ای است که به شکل «نامه‌ها» نیز نوشته می‌شود.

کدهای نمایشی در این توصیف به زبان پیتون هستند. این کدها فقط باید برای بهتر فهمیده‌شدن الگوریتم استفاده شوند و نباید به شکل مستقیم در برنامه‌ها استفاده شوند. (مشخصاً، مجوز این کدها، که مجوز مستندات آزاد گنو است، اجازه نمی‌دهد که این کدها بدون اجازهٔ مالک حقوق توصیف حاضر در هیچ برنامه‌ای استفاده شوند.) به‌علاوه، این کدها به منظور کمک به بهتر فهمیده‌شدن، ساده شده‌اند و ممکن است موارد خاص را در نظر نگرفته باشند و/یا برای پیاده‌سازی سریع مناسب نباشند. برای اطمینان از درستی پیاده‌سازیهای این توصیف، نباید از این کدها استفاده کرد، بلکه باید از داده‌های آزمون ضمیمهٔ این توصیف استفاده کرد.

۶ ترتیب‌بندی دو دنبالهٔ عددی

الگوریتم زیر برای ترتیب‌بندی دو دنبالهٔ عددی به کار می‌رود.

ورودی: دو دنبالهٔ عددی a (به طول l) و b (به طول m).

خروجی: حاصل مقایسه («قبل»، «بعد»، یا «هم‌ارز»).

۱) گام ۲ را برای هر i از ۰ تا $\min(l, m)$-۱، اجرا کنید. سپس به گام ۳ بروید.

۲) اگر a_i کمتر از b_i است (به این فرض که رشته از موقعیت ۰ شروع می‌شود)، نتیجهٔ مقایسه «قبل» است و الگوریتم را تمام کنید. اگر b_i کمتر از a_i است، نتیجهٔ مقایسه «بعد» است و الگوریتم را تمام کنید.

۳) اگر l کمتر از m است، نتیجهٔ مقایسه «قبل» است و الگوریتم را تمام کنید. اگر m کمتر از l است، نتیجهٔ مقایسه «بعد» است و الگوریتم را تمام کنید. اگر l و m برابر هستند نتیجهٔ مقایسه «هم‌ارز» است. الگوریتم را تمام کنید.

کد نمایشی:

```
i = 0
while i < l and i < m:
    if a[i] < b[i]:
        return -1
    elif a[i] > b[i]:
        return +1
    i += 1
if l < m:
    return -1
elif l > m:
    return +1
else:
    return 0
```

۷ پیش‌پردازش رشته‌های فارسی

در این توصیف فرض می‌شود که رشته‌های ورودی به صورت مطلوب جهت پردازش برای ترتیب‌بندی فارسی هستند. به منظور تبدیل رشته‌ها به صورت مطلوب، باید پیش‌پردازش زیر روی آنها انجام شود:

ورودی: رشتهٔ s جهت پیش‌پردازش.

خروجی: رشتهٔ پیش‌پردازش‌شدهٔ s.

۱) اگر از نویسه‌های بلوکهای «شکلهای نمایشی عربی» یونی‌کد (با نام انگلیسی Arabic Presentation Forms) یعنی نویسه‌های محدودهٔ U+FB50..U+FDFF (به جز U+FD3E و U+FD3F) و U+FE80 تا U+FEFE در رشتهٔ s استفاده شده بود، آن قسمتهای رشته را با حفظ کامل معنا و در نظر گرفتن احتمال نیاز به درج نویسه‌های فاصلهٔ مجازی و اتصال مجازی، به رشته‌ای که از نویسه‌های عادی فارسی، یعنی U+0600..U+06FF، استفاده می‌کند تبدیل کنید. (مثلاً رشتهٔ <ع آخر، م اول> («عم») به رشتهٔ <اتصال مجازی، ع، فاصلهٔ مجازی، م، اتصال مجازی> و رشتهٔ «لآ» به رشتهٔ <اتصال مجازی، ل، آ> تبدیل می‌شود.)

۲) نویسه‌های فاصلهٔ مجازی و اتصال مجازی موجود در رشتهٔ s را در صورتی که زائد بودند، یعنی حذف آنها در هیچ شرایطی در شیوهٔ نمایش رشته‌ها تغییری ایجاد نمی‌کرد، حذف کنید. (مثلاً رشتهٔ <ل، ک، فاصلهٔ مجازی، فاصلهٔ مجازی، ل، ک> («لک‌لک») به رشتهٔ <ل، ک، فاصلهٔ مجازی، ل، ک> تبدیل می‌شود چون یکی از نویسه‌های فاصلهٔ مجازی زائد است و در هیچ شرایطی تغییری در نمایش رشته ایجاد نمی‌کند.)

۳) رشتهٔ s را به صورت نرمال C (NFC)، مشخص شده در ضمیمهٔ ۱۵ استاندارد یونی‌کد، تبدیل کنید. (مثلاً رشتهٔ <ر، الف، ـٔ، ی> به رشتهٔ <ر، أ، ی> تبدیل می‌شود.)

۸ ترتیب‌بندی دو رشتهٔ فارسی

در الگوریتم زیر فرض می‌شود که دو رشتهٔ ورودی محدود به نویسه‌های فارسی، عربی، و مشترکی هستند که در بخش «فهرست نویسه‌های پشتیبانی‌شده» این توصیف آمده‌اند. این الگوریتم نمی‌تواند رشته‌های دیگر را ترتیب‌بندی کند. رشته‌هایی که شامل نویسه‌های دیگری نیز باشند باید بر اساس استاندارد بین‌المللی ISO/IEC 14651 و استاندارد فنی شمارهٔ ۱۰ یونی‌کد، ولی با سازگاری کامل با ویژگیهای مشخص‌شده در این توصیف برای نویسه‌های پشتیبانی‌شده، مرتب شوند.

در این الگوریتم سطوح مختلف ترتیب‌بندی به ترتیب اجرا می‌شوند و تنها در صورتی که دو رشته در یک سطح هم‌ارز بودند، سطح بعدی بررسی می‌شود.

کاربردها می‌توانند بسته به نیاز، سطوح دیگری به انتهای فهرست اضافه کنند تا در صورت نیاز بین رشته‌هایی که الگوریتم زیر آنها را «هم‌ارز» تشخیص می‌دهد ترتیب قائل شوند.

ورودی: دو رشتهٔ پیش‌پردازش‌شده و فارسی s و t.

خروجی: حاصل مقایسه («قبل»، «بعد»، یا «هم‌ارز»).

۱) دو رشتهٔ s و t را بر اساس الگوریتم «پیش‌پردازش رشته‌های فارسی» پردازش کنید. اجرای این گام در صورتی که بتوان فرض کرد رشته‌های ورودی از قبل برای پردازش توسط الگوریتم حاضر مناسبند، لزومی ندارد.

۲) به ازای هر i از ۱ تا ۳ گام ۳ را اجرا کنید. سپس به گام ۴ بروید.

۳) دو رشتهٔ s و t را بر اساس الگوریتم «پردازش سطح i» پردازش کنید تا دنباله‌های عددی a و b به دست آیند. دو رشتهٔ a و b را بر اساس الگوریتم «ترتیب‌بندی دو دنبالهٔ عددی» ترتیب‌بندی کنید. اگر خروجی الگوریتم ترتیب‌بندی «قبل» یا «بعد» بود، نتیجهٔ مقایسه همان خروجی است و الگوریتم را تمام کنید.

۴) نتیجهٔ مقایسه هم ارز است. الگوریتم را تمام کنید.

کد نمایشی:

```
s = prep (s)
t = prep (t)
for i in [1, 2, 3]:
    a = weightlist (s, i)
    b = weightlist (t, i)
    r = numcollate (a, b)
    if (r != 0):
        return r
return 0
```

۹ پردازش سطح یک، نویسه‌های اصلی

در این سطح، نویسه‌های اصلی فارسی مقایسه می‌شوند، به این معنی که فرض می‌شود نویسه‌های فرعی در رشته وجود ندارند و نویسه‌های اصلی مشابه، معادل در نظر گرفته می‌شوند. تنها در صورتی که دو رشته در این سطح هم‌ارز بودند، سطوح دیگر بررسی می‌شوند. الگوریتم مشخص شده در این بخش و بخشهای مشابه، عملاً یک رشته را به یک دنبالهٔ عددی تبدیل می‌کند تا حاصل ترتیب‌بندی دو رشته، همان حاصل ترتیب‌بندی دو دنبالهٔ عددی باشد.

ورودی: رشتهٔ پیش‌پردازش‌شده و فارسی s.

خروجی: دنبالهٔ عددی a جهت استفاده در ترتیب‌بندی.

۱) دنبالهٔ عددی a را خالی کنید.

۲) به ازای هر نویسهٔ c موجود در رشتهٔ s، گام ۳ را اجرا کنید. سپس الگوریتم را تمام کنید.

۳) اگر نویسهٔ c در جدول ۱ آمده است، وزن متناظرش را به انتهای دنبالهٔ a اضافه کنید.

کد نمایشی:

```
a = []
for c in s:
    if table1.haskey (c):
        a.append (table1 [c])
return a
```

مثال: خروجی الگوریتم فوق برای رشتهٔ <م، ز، د، ـَ، ك> («مزدَك» با کاف عربی)، دنبالهٔ [۴۰، ۲۵، ۲۲، ۳۷] است.

جدول ۱: وزن نویسه‌ها برای سطح یک

نام نویسه	شکل نویسه	کد یونی‌کد	وزن
رقم فارسی صفر	۰	U+06F0	۱
رقم صفر عربی	۰	U+0660	۱
رقم فارسی یک	۱	U+06F1	۲
رقم یک عربی	۱	U+0661	۲

جدول ۱: وزن نویسه‌ها برای سطح یک			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
رقم فارسی دو	۲	U+06F2	۳
رقم دو عربی	٢	U+0662	۳
رقم فارسی سه	۳	U+06F3	۴
رقم سه عربی	٣	U+0663	۴
رقم فارسی چهار	۴	U+06F4	۵
رقم چهار عربی	٤	U+0664	۵
رقم فارسی پنج	۵	U+06F5	۶
رقم پنج عربی	٥	U+0665	۶
رقم فارسی شش	۶	U+06F6	۷
رقم شش عربی	٦	U+0666	۷
رقم فارسی هفت	۷	U+06F7	۸
رقم هفت عربی	٧	U+0667	۸
رقم فارسی هشت	۸	U+06F8	۹
رقم هشت عربی	٨	U+0668	۹

جدول ۱: وزن نویسه‌ها برای سطح یک			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
رقم فارسی نه	۹	U+06F9	۱۰
رقم نه عربی	۹	U+0669	۱۰
حرف فارسی آ	آ	U+0622	۱۱
حرف فارسی الف	ا	U+0627	۱۲
حرف الف وصل	ٱ	U+0671	۱۲
حرف فارسی همزه	ء	U+0628	۱۳
حرف فارسی الف با همزه‌ی بالا	أ	U+0623	۱۳
حرف الف با همزه‌ی پایین	إ	U+0625	۱۳
حرف فارسی واو با همزه‌ی بالا	ؤ	U+0624	۱۳
حرف فارسی ی با همزه‌ی بالا	ئ	U+0626	۱۳
حرف فارسی ب	ب	U+062B	۱۴
حرف فارسی پ	پ	U+067E	۱۵
حرف فارسی ت	ت	U+062A	۱۶
حرف فارسی ث	ث	U+062B	۱۷

جدول ۱: وزن نویسه‌ها برای سطح یک			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
حرف فارسی جیم	ج	U+062C	۱۸
حرف فارسی چ	چ	U+0686	۱۹
حرف فارسی ح	ح	U+062D	۲۰
حرف فارسی خ	خ	U+062E	۲۱
حرف فارسی دال	د	U+062F	۲۲
حرف فارسی ذال	ذ	U+0630	۲۳
حرف فارسی ر	ر	U+0631	۲۴
حرف فارسی ز	ز	U+0632	۲۵
حرف فارسی ژ	ژ	U+0698	۲۶
حرف فارسی سین	س	U+0633	۲۷
حرف فارسی شین	ش	U+0634	۲۸
حرف فارسی صاد	ص	U+0635	۲۹
حرف فارسی ضاد	ض	U+0636	۳۰
حرف فارسی طا	ط	U+0637	۳۱

جدول ۱: وزن نویسه‌ها برای سطح یک			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
حرف فارسی ظا	ظ	U+0638	۳۲
حرف فارسی عین	ع	U+0639	۳۳
حرف فارسی غین	غ	U+063A	۳۴
حرف فارسی ف	ف	U+0641	۳۵
حرف فارسی قاف	ق	U+0642	۳۶
حرف فارسی کاف	ک	U+06A9	۳۷
حرف کاف عربی	ك	U+0643	۳۷
حرف فارسی گاف	گ	U+06AF	۳۸
حرف فارسی لام	ل	U+0644	۳۹
حرف فارسی میم	م	U+0645	۴۰
حرف فارسی نون	ن	U+0646	۴۱
حرف فارسی واو	و	U+0648	۴۲
حرف فارسی ه	ه	U+0647	۴۳
حرف ت گرد	ة	U+0629	۴۳

جدول ۱: وزن نویسه‌ها برای سطح یک			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
حرف فارسی ی	ی	U+06CC	۴۴
حرف ی عربی بی‌نقطه	د	U+0649	۴۴
حرف ی عربی نقطه‌دار	ي	U+064A	۴۴

۱۰ پردازش سطح دو، تفاوتهای فرعی

در این سطح، تفاوتهای فرعی نویسه‌های اصلی فارسی در نظر گرفته می‌شوند، به این معنی که فرض می‌شود نویسه‌های فرعی در این سطح هنوز در رشته وجود ندارند و فقط تفاوت نویسه‌های اصلی در نظر گرفته می‌شود، مثلاً تمایز بین انواع همزه‌ها یا شکل فارسی و عربی نویسه‌ها.

تنها در صورتی که دو رشته در سطح یک هم‌ارز بودند این سطح بررسی می‌شود. تنها در صورتی که دو رشته در این سطح هم‌ارز بودند، سطح بعدی بررسی می‌شود.

الگوریتم مشخص شده در این بخش و بخشهای مشابه، عملاً یک رشته را به یک دنبالهٔ عددی تبدیل می‌کند تا حاصل ترتیب‌بندی دو رشته، همان حاصل ترتیب‌بندی دو دنبالهٔ عددی باشد.

باید توجه شود که این الگوریتم همان الگوریتم پردازش سطح یک است، که فقط وزن نویسه‌ها در آن متفاوت است.

ورودی: رشتهٔ پیش‌پردازش‌شده و فارسی s.

خروجی: دنبالهٔ عددی a جهت استفاده در ترتیب‌بندی.

۱) دنبالهٔ عددی a را خالی کنید.

۲) به ازای هر نویسهٔ c موجود در رشتهٔ s، گام ۳ را اجرا کنید. سپس الگوریتم را تمام کنید.

۳) اگر نویسهٔ c در جدول ۲ آمده است، وزن متناظرش را به انتهای دنبالهٔ a اضافه کنید.

کد نمایشی:

```
a = []
for c in s:
    if table2.haskey (c):
        a.append (table2 [c])
return a
```

مثال: خروجی الگوریتم فوق برای رشتهٔ <م، ز، د، ـَ، ك> («مزدَك» با کاف عربی)، دنبالهٔ [۱، ۱، ۱، ۱۰] است.

جدول ۲: وزن نویسه‌ها برای سطح دو			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
حرف فارسی آ	آ	U+0622	۱
حرف فارسی الف	ا	U+0627	۱
حرف الف وصل	ٱ	U+0671	۴

جدول ۲: وزن نویسه‌ها برای سطح دو			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
حرف فارسی همزه	ء	U+0628	۱
حرف فارسی الف با همزه‌ی بالا	أ	U+0623	۲
حرف الف با همزه‌ی پایین	إ	U+0625	۳
حرف فارسی واو با همزه‌ی بالا	ؤ	U+0624	۵
حرف فارسی ی با همزه‌ی بالا	ئ	U+0626	۶
حرف فارسی ب	ب	U+062B	۱
حرف فارسی پ	پ	U+067E	۱
حرف فارسی ت	ت	U+062A	۱
حرف فارسی ث	ث	U+062B	۱
حرف فارسی جیم	ج	U+062C	۱
حرف فارسی چ	چ	U+0686	۱
حرف فارسی ح	ح	U+062D	۱
حرف فارسی خ	خ	U+062E	۱
حرف فارسی دال	د	U+062F	۱

جدول ۲: وزن نویسه‌ها برای سطح دو			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
حرف فارسی ذال	ذ	U+0630	۱
حرف فارسی ر	ر	U+0631	۱
حرف فارسی ز	ز	U+0632	۱
حرف فارسی ژ	ژ	U+0698	۱
حرف فارسی سین	س	U+0633	۱
حرف فارسی شین	ش	U+0634	۱
حرف فارسی صاد	ص	U+0635	۱
حرف فارسی ضاد	ض	U+0636	۱
حرف فارسی طا	ط	U+0637	۱
حرف فارسی ظا	ظ	U+0638	۱
حرف فارسی عین	ع	U+0639	۱
حرف فارسی غین	غ	U+063A	۱
حرف فارسی ف	ف	U+0641	۱
حرف فارسی قاف	ق	U+0642	۱

جدول ۲: وزن نویسه‌ها برای سطح دو			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
حرف فارسی کاف	ک	U+06A9	۱
حرف کاف عربی	ك	U+0643	۱۰
حرف فارسی گاف	گ	U+06AF	۱
حرف فارسی لام	ل	U+0644	۱
حرف فارسی میم	م	U+0645	۱
حرف فارسی نون	ن	U+0646	۱
حرف فارسی واو	و	U+0648	۱
حرف فارسی ه	ه	U+0647	۱
حرف ت گرد	ة	U+0629	۸
حرف فارسی ی	ی	U+06CC	۱
حرف ی عربی بی‌نقطه	د	U+0649	۷
حرف ی عربی نقطه‌دار	ي	U+064A	۹
رقم فارسی صفر	۰	U+06F0	۱
رقم صفر عربی	٠	U+0660	۱۰

جدول ۲: وزن نویسه‌ها برای سطح دو			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
رقم فارسی یک	۱	U+06F1	۱
رقم یک عربی	١	U+0661	۱۰
رقم فارسی دو	۲	U+06F2	۱
رقم دو عربی	٢	U+0662	۱۰
رقم فارسی سه	۳	U+06F3	۱
رقم سه عربی	٣	U+0663	۱۰
رقم فارسی چهار	۴	U+06F4	۱
رقم چهار عربی	٤	U+0664	۱۰
رقم فارسی پنج	۵	U+06F5	۱
رقم پنج عربی	٥	U+0665	۱۰
رقم فارسی شش	۶	U+06F6	۱
رقم شش عربی	٦	U+0666	۱۰
رقم فارسی هفت	۷	U+06F7	۱
رقم هفت عربی	٧	U+0667	۱۰

جدول ۲: وزن نویسه‌ها برای سطح دو			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
رقم فارسی هشت	۸	U+06F8	۱
رقم هشت عربی	۸	U+0668	۱۰
رقم فارسی نه	۹	U+06F9	۱
رقم نه عربی	۹	U+0669	۱۰

۱۱ پردازش سطح سه، نویسه‌های فرعی

در این سطح، نویسه‌های فرعی فارسی در نظر گرفته می‌شوند، به این معنی که تقریباً فرض می‌شود نویسه‌های اصلی در این سطح در رشته وجود ندارند و فقط تفاوت نویسه‌های فرعی در نظر گرفته می‌شود، مثلاً فاصله‌ها، اِعرابها، و علائم نقطه‌گذاری.

تنها در صورتی که دو رشته در سطح یک و دو هم‌ارز بودند این سطح بررسی می‌شود.

الگوریتم مشخص شده در این بخش و بخشهای مشابه، عملاً یک رشته را به یک دنبالهٔ عددی تبدیل می‌کند تا حاصل ترتیب‌بندی دو رشته، همان حاصل ترتیب‌بندی دو دنبالهٔ عددی باشد.

توجه کنید که این الگوریتم با دو الگوریتم سطح یک و دو متفاوت است، از این نظر که دنبالهٔ حاصل از آن شامل موقعیت نویسه‌های فرعی در رشتهٔ اصلی نیز هست. این دنبالهٔ موقعیتها برای تعیین ترتیب رشته‌ها از جمله در مواردی که اعرابهای دو رشته یکسان است و فقط پایه‌شان تفاوت دارد مفید است.

ورودی: رشتهٔ پیش‌پردازش‌شده و فارسی s.

خروجی: دنبالهٔ عددی a جهت استفاده در ترتیب‌بندی.

۱) دنباله‌های عددی a و il را خالی کنید.

۲) به ازای هر نویسهٔ c موجود در رشتهٔ s که در موقعیت iام رشته قرار دارد (فرض می‌شود که رشته از موقعیت ۰ شروع می‌شود)، گام ۳ را اجرا کنید. سپس به گام ۴ بروید.

۳) اگر نویسهٔ c در جدول ۳ آمده است، وزن متناظرش را به انتهای دنبالهٔ a اضافه کنید و عدد $i+1$ را نیز به انتهای دنبالهٔ il اضافه کنید.

۴) عدد ۰ را به انتهای دنبالهٔ a اضافه کنید. دنبالهٔ il را به انتهای دنبالهٔ a ضمیمه کنید. سپس الگوریتم را تمام کنید.

کد نمایشی:

```
a = il = []
i = 0
while i < len (s):
    c = s [i]
    if table3.haskey (c):
        a.append (table3 [c])
        il.append (i+1)
    i += 1
return a+[0]+il
```

مثال: خروجی الگوریتم فوق برای رشتهٔ <م، ز، د، ـَ ك> («مزدَك» با کاف عربی)، دنبالهٔ [۴، ۰، ۴] است.

خروجی الگوریتم فوق برای رشتهٔ <ع، ـَ ا، ل، ـِ م> («عَالِم»)، دنبالهٔ [۴، ۵، ۰، ۲، ۵] و برای رشتهٔ <ع، ـَ ا، ل، م، ـِ> («عَالمِ»)، دنبالهٔ [۴، ۵، ۰، ۲، ۶] است. این مسئله باعث می‌شود رشتهٔ اول قبل از رشتهٔ دوم ترتیب‌بندی شود.

جدول ۳: وزن نویسه‌ها برای سطح سه			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
فاصله		U+0020	۱
فاصلهٔ مجازی		U+200C	۲
اتصال مجازی		U+200D	۳
زبر فارسی (فتحه)	ـَ	U+064E	۴
زیر فارسی (کسره)	ـِ	U+0650	۵
پیش فارسی (ضمه)	ـُ	U+064F	۶
دوزبر فارسی (تنوین نصب)	ـً	U+064B	۷
دوزیر فارسی (تنوین جر)	ـٍ	U+064D	۸
دوپیش فارسی (تنوین رفع)	ـٌ	U+064C	۹
تشدید فارسی	ـّ	U+0651	۱۰
ساکن فارسی	ـْ	U+0652	۱۱
مد فارسی	ـٓ	U+0653	۱۲
همزه‌ی فارسی بالا	ـٔ	U+0654	۱۳
همزه‌ی فارسی پایین	ـٕ	U+0655	۱۴

جدول ۳: وزن نویسه‌ها برای سطح سه			
نام نویسه	شکل نویسه	کد یونی‌کد	وزن
الف مقصوره‌ی فارسی	ـٰ	U+0670	۱۵
الف مقصوره‌ی فارسی پایین	ٰ	U+0647	۱۶

۱۲ نویسه‌های پشتیبانی‌شده

نویسه‌های زیر در این توصیف پشتیبانی می‌شوند، به این معنی که اگر همهٔ نویسه‌های دو رشته‌ای که باید ترتیب‌بندی شوند از فهرست زیر باشند، توصیف حاضر می‌تواند ترتیب آنها را تعیین کند. در غیر این صورت، یعنی اگر حداقل یکی از رشته‌ها حداقل یک نویسه داشته باشد که در این بخش ذکر نشده است، رشته‌ها باید، با در نظر گرفتن ترتیب‌بندی توصیف حاضر، بر اساس استاندارد بین‌المللی ISO/IEC 14651 و/یا استاندارد فنی شمارهٔ ۱۰ یونی‌کد ترتیب‌بندی شوند.

فهرست نویسه‌ها به شرح زیر است:

جدول ۴: نویسه‌های پشتیبانی‌شده

نام نویسه	شکل نویسه	کد یونی‌کد
حرف فارسی آ	آ	U+0622
حرف فارسی الف	ا	U+0627
حرف الف وصل	ٱ	U+0671
حرف فارسی همزه	ء	U+0628
حرف فارسی الف با همزه‌ی بالا	أ	U+0623
حرف الف با همزه‌ی پایین	إ	U+0625
حرف فارسی واو با همزه‌ی بالا	ؤ	U+0624
حرف فارسی ی با همزه‌ی بالا	ئ	U+0626
حرف فارسی ب	ب	U+062B
حرف فارسی پ	پ	U+067E

جدول ۴: نویسه‌های پشتیبانی‌شده		
نام نویسه	شکل نویسه	کد یونی‌کد
حرف فارسی ت	ت	U+062A
حرف فارسی ث	ث	U+062B
حرف فارسی جیم	ج	U+062C
حرف فارسی چ	چ	U+0686
حرف فارسی ح	ح	U+062D
حرف فارسی خ	خ	U+062E
حرف فارسی دال	د	U+062F
حرف فارسی ذال	ذ	U+0630
حرف فارسی ر	ر	U+0631
حرف فارسی ز	ز	U+0632
حرف فارسی ژ	ژ	U+0698
حرف فارسی سین	س	U+0633
حرف فارسی شین	ش	U+0634
حرف فارسی صاد	ص	U+0635

جدول ۴: نویسه‌های پشتیبانی‌شده		
نام نویسه	شکل نویسه	کد یونی‌کد
حرف فارسی ضاد	ض	U+0636
حرف فارسی طا	ط	U+0637
حرف فارسی ظا	ظ	U+0638
حرف فارسی عین	ع	U+0639
حرف فارسی غین	غ	U+063A
حرف فارسی ف	ف	U+0641
حرف فارسی قاف	ق	U+0642
حرف فارسی کاف	ک	U+06A9
حرف کاف عربی	ك	U+0643
حرف فارسی گاف	گ	U+06AF
حرف فارسی لام	ل	U+0644
حرف فارسی میم	م	U+0645
حرف فارسی نون	ن	U+0646
حرف فارسی واو	و	U+0648

جدول ۴: نویسه‌های پشتیبانی‌شده		
نام نویسه	شکل نویسه	کد یونی‌کد
حرف فارسی ه	ه	U+0647
حرف ت گرد	ة	U+0629
حرف فارسی ی	ی	U+06CC
حرف ی عربی بی‌نقطه	د	U+0649
حرف ی عربی نقطه‌دار	ي	U+064A
رقم فارسی صفر	۰	U+06F0
رقم صفر عربی	٠	U+0660
رقم فارسی یک	۱	U+06F1
رقم یک عربی	١	U+0661
رقم فارسی دو	۲	U+06F2
رقم دو عربی	٢	U+0662
رقم فارسی سه	۳	U+06F3
رقم سه عربی	٣	U+0663
رقم فارسی چهار	۴	U+06F4

جدول ۴: نویسه‌های پشتیبانی‌شده		
نام نویسه	شکل نویسه	کد یونی‌کد
رقم چهار عربی	٤	U+0664
رقم فارسی پنج	۵	U+06F5
رقم پنج عربی	٥	U+0665
رقم فارسی شش	۶	U+06F6
رقم شش عربی	٦	U+0666
رقم فارسی هفت	۷	U+06F7
رقم هفت عربی	٧	U+0667
رقم فارسی هشت	۸	U+06F8
رقم هشت عربی	٨	U+0668
رقم فارسی نه	۹	U+06F9
رقم نه عربی	٩	U+0669
فاصله		U+0020
فاصلهٔ مجازی		U+200C
اتصال مجازی		U+200D

جدول ۴: نویسه‌های پشتیبانی‌شده		
نام نویسه	شکل نویسه	کد یونی‌کد
زبر فارسی (فتحه)	ـَ	U+064E
زیر فارسی (کسره)	ـِ	U+0650
پیش فارسی (ضمه)	ـُ	U+064F
دوزبر فارسی (تنوین نصب)	ـً	U+064B
دوزیر فارسی (تنوین جر)	ـٍ	U+064D
دوپیش فارسی (تنوین رفع)	ـٌ	U+064C
تشدید فارسی	ـّ	U+0651
ساکن فارسی	ـْ	U+0652
مد فارسی	ـٓ	U+0653
همزه‌ی فارسی بالا	ـٔ	U+0654
همزه‌ی فارسی پایین	ـٕ	U+0655
الف مقصوره‌ی فارسی	ـٰ	U+0670
الف مقصوره‌ی فارسی پایین	ٰ	U+0647
نشانه‌ی چپ‌به‌راست		U+200E

جدول ۴: نویسه‌های پشتیبانی‌شده		
نام نویسه	**شکل نویسه**	**کد یونی‌کد**
نشانه‌ی راست‌به‌چپ		U+200F
زیرمتن چپ‌به‌راست		U+202A
زیرمتن راست‌به‌چپ		U+202B
پایان زیرمتن		U+202C
زیرمتن اکیداً چپ‌به‌راست		U+202D
زیرمتن اکیداً چپ‌به‌راست		U+202E
نشانه‌ی ترتیب بایت‌ها		U+FEFF

۱۳ مراجع

[1] ISO/IEC 10646:2003, “Information technology — Universal Multiple-Octet Coded Character Set (UCS)”.

[2] ISO/IEC 14651:2001, “Information technology — International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering”.

[3] ISO/IEC TR 14652:2002, “Information technology — Specification method for cultural conventions”.

[4] The Unicode Consortium, *The Unicode Standard, Version 4.0*, Reading, Massachusetts, Addison-Wesley, 2003. ISBN 0-321-18578-1.

[5] Mark Davis, Unicode Standard Annex #9, “The Bidirectional Algorithm”, Version 4.0.1, 2004.

[6] Mark Davis and Ken Whistler, Unicode Technical Standard #10, “Unicode Collation Algorithm”, Version 4.0, 2004.

[7] Mark Davis and Martin Dürst, Unicode Standard Annex #15, “Unicode Normalization Forms”, Version 4.0.0, 2003.

[8] Mark Davis, Unicode Technical Standard #35, “Locale Data Markup Language (LDML)”, Version 1.1 (draft), 2004.

[9] Donald Knuth, *The Art of Computer Programming, Volume 3: Sorting and Searching*, Second Edition, Reading, Massachusetts, Addison-Wesley, 1998. ISBN 0-201-89685-0.

[10] *Wikipedia: The Free Encyclopedia*, http://en.wikipedia.org/.

[۱۱] استاندارد ملی شمارهٔ ۳۳۴۲ سال ۱۳۷۲، «استاندارد کد تبادل اطلاعات ۸ بیتی فارسی».

[۱۲] استاندارد ملی شمارهٔ ۶۲۱۹ سال ۱۳۸۱، «فنّاوریِ اطلاعات ــ تبادل و شیوه‌ی نمایشِ اطلاعاتِ فارسی بر اساسِ یونی‌کُد».

[۱۳] *دستور خط فارسی، مصوب فرهنگستان زبان و ادب فارسی*، فرهنگستان زبان و ادب فارسی (نشر آثار)، تهران، ۱۳۸۱. شابک ۳-۱۳-۷۵۳۱-۹۶۴.

[۱۴] *شیوه‌نامه*، مرکز نشر دانشگاهی، ویرایش دوم، تهران، ۱۳۷۲. شابک ۷-۸۱۲۷-۰۱-۹۶۴.

[۱۵] *فرهنگ فارسی معین*؟؟؟؟

[۱۶] غلامحسین صدری افشار، نسرین حکمی، و نسترن حکمی، *فرهنگ معاصر فارسی*، فرهنگ معاصر، تهران، ۱۳۸۱. شابک ۰-۷۳-۵۵۴۵-۹۶۴.

[۱۷] حسن عمید، *فرهنگ فارسی عمید*، مؤسسهٔ انتشارات امیرکبیر، تهران، ۱۳۷۳. شابک ۷-۰۱۳۱-۰۰-۹۶۴.

[۱۸] غلامحسین مصاحب و دیگران، *دایرةالمعارف فارسی*، شرکت سهامی کتابهای جیبی، تهران، ۱۳۸۱. شابک X- ۰۴۴-۳۰۳-۹۶۴.

[۱۹] *فرهنگ فارسی سخن*؟؟؟؟

[۲۰] علی‌اکبر دهخدا و دیگران، *لغت‌نامه*، مؤسسهٔ انتشارات و چاپ دانشگاه تهران، تهران، ۱۳۷۲. شابک ۴-۰۰۰۰-۰۳-۹۶۴.

[۲۱] محمدرضا محمدی‌فر، *مبانی نمایه‌سازی*، سازمان چاپ و انتشارات وزارت فرهنگ و ارشاد اسلامی، تهران، ۱۳۸۱.

[۲۲] میرشمس‌الدین ادیب سلطانی، *راهنمای آماده‌سازی کتاب*، ویراست سوم، شرکت انتشارات علمی و فرهنگی، تهران، ۱۳۸۱. شابک ۰-۳۳۶-۴۴۵-۹۶۴.

۱۴ اجازه‌نامهٔ مستندات آزاد گنو (GNU Free Documentation License)

Version 1.2, November 2002

Copyright © 2000,2001,2002 Free Software Foundation, Inc.
59 Temple Place, Suite 330, Boston, MA 02111-1307 USA
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the

subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to

be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- **A.** Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- **B.** List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- **C.** State on the Title page the name of the publisher of the Modified Version, as the publisher.
- **D.** Preserve all the copyright notices of the Document.
- **E.** Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- **F.** Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- **G.** Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- **H.** Include an unaltered copy of this License.
- **I.** Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- **J.** Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- **K.** For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- **L.** Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- **M.** Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

- **N.** Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- **O.** Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements."

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your

rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (c)  YEAR  YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.2
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover
Texts.  A copy of the license is included in the section entitled
"GNU Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being
LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.