

# تشخیص خودکار وزن عروضی اشعار کلاسیک فارسی\*

وحید موجی<sup>۱</sup> و محرم اسلامی<sup>۲</sup>

<sup>۱</sup> دانشگاه صنعتی شریف

<sup>۲</sup> دانشگاه زنجان

۲۵ اردیبهشت ۱۳۹۲

## چکیده

در مقاله حاضر می‌کوشیم وزن عروضی اشعار کلاسیک فارسی را به طور خودکار و ماشینی تعیین کنیم. برخورداری خط فارسی از ویژگی‌هایی چون عدم نمایش واژه‌های کوتاه و به دنبال آن موضوع هم‌نویسی، مسأله کسره اضافه، فاصله بین اجزای کلمه واحد، فقدان فاصله بین کلمه‌های مستقل، موضوع جدا نویسی و پیوسته نویسی و غیره موجب می‌شود قبل از انجام هرگونه پردازشی، متون فارسی را به زنجیره واجی تبدیل کنیم. بدین منظور با استفاده از تحلیل گر صرفی پارس-مورف که توسط نگارندگان طراحی و پیاده‌سازی شده است، متن ورودی از لحاظ صرفی تحلیل شده و اجزای صرفی آن از قبیل پیشوندها، پسوندها، اشتقاق و ترکیب بدست آمده و سپس با استفاده از واژگان زبانی فارسی (اسلامی و دیگران، ۱۳۸۳)، صورت واجی آنها با هم ترکیب شده و در نهایت صورت واجی متن ورودی به دست می‌آید. سپس این زنجیره واجی را طبق روش‌های موجود در عروض فارسی تقطیع هجایی و تقطیع به ارکان کرده و در ادامه با استفاده از الگوریتم لونشتاین که الگوریتمی برای یافتن حداقل فاصله بین دو رشته است، در بین گروه‌های وزنی که گردآوری کرده‌ایم، آن گروهی که کمترین فاصله را با تقطیع متن مورد نظر دارد انتخاب کرده و وزن مناسب را مطابق با آن می‌یابیم. از آنجا که خط فارسی فاقد مصوت‌ها است، لذا دقت الگوریتم ممکن است بسته به نمود واجی که پیشنهاد می‌دهد تغییر کند یا کاهش یابد. به همین دلیل این امکان در برنامه فراهم شده است که شعر را به صورت زنجیره واجی نیز وارد نموده و وزن آن را تعیین کنیم. در این کار از ۳۱ وزن رایج و پرکاربرد زبان فارسی استفاده کرده‌ایم. از کاربردهای برنامه تشخیص خودکار وزن عروضی می‌توان به آموزش عروض به فارسی‌زبانان و غیر فارسی‌زبانان، بررسی صحت و دقت تعیین وزن عروضی اشعار به روش‌های سنتی و ارائه شیوه‌های مختلف خوانش اشعار کلاسیک فارسی اشاره نمود. کلیدواژه‌ها: تحلیل گر صرفی، زنجیره واجی، تشخیص خودکار، وزن عروضی، الگوریتم لونشتاین، شعر فارسی.

## ۱ مقدمه

در این پژوهش سعی بر این داریم تا با استفاده از فناوری رایانه و روش‌های زبان‌شناسی رایانشی، وزن شعر کلاسیک فارسی را تعیین کنیم. از آنجا که در شعر کلاسیک، وزن مصراع معیار است، لذا ورودی این برنامه «مصراع» می‌باشد نه «بیت». یکی از نیازمندی‌های چنین برنامه‌ای داشتن قابلیت تبدیل متن فارسی به زنجیره

\*دومین همایش ملی آموزش زبان فارسی و زبان‌شناسی، شیراز، اردیبهشت ۱۳۹۲

واجی می‌باشد. روش‌های متعددی برای تبدیل متن به زنجیره واجی مورد استفاده قرار گرفته است. در Allen (1992) از قواعد تبدیل نویسه به صورت واجی استفاده شده و استثنائات نیز از یک فرهنگ استخراج می‌شود. استفاده از یک درخت تصمیم چندسطحی که هر نویسه را نسبت به حروف مجاور آن به صورت یک درخت نمایش می‌دهد در Torkkola (1993) مورد مطالعه قرار گرفته است. استفاده از روش‌های زبان طبیعی نیز مورد بررسی قرار گرفته است. در این روش، هر تکواژ به همراه اطلاعات مربوط به صورت‌های صرفی مختلف آن مانند صورت جمع، گذشته، حال، و غیره و کلیه اطلاعات صرفی مربوطه در یک دادگان ذخیره می‌شود. در این حالت نیز اگر کلمه در فهرست تکواژها نباشد از قواعد تبدیل نویسه به صورت واجی یا فرهنگ استنهاها استفاده می‌شود (صیادیان و نصیرزاده، ۱۳۷۵).

یکی از روش‌های تبدیل خودکار متن فارسی به زنجیره واجی، استفاده از یک تحلیل‌گر صرفی است. در این روش با استفاده از یک تحلیل‌گر صرفی، عبارت ورودی به اجزای تشکیل‌دهنده آن تقطیع شده و سپس هر قطعه به عنوان ورودی به تحلیل‌گر صرفی داده می‌شود تا به اجزای صرفی تشکیل‌دهنده‌اش از قبیل پسوندها و پیشوندهای تصریفی و اشتقاقی، پایه‌های ترکیب و اشتقاق و غیره تجزیه شود و سپس صورت واجی اجزا با هم ترکیب شده و زنجیره واجی کل متن به دست می‌آید.

در این مقاله از روش تحلیل‌گر صرفی برای تبدیل متن فارسی به زنجیره واجی استفاده شده است. سپس زنجیره واجی تقطیع هجایی و تقطیع به ارکان شده و به صورت رشته‌ای از (-ها و (U)ها درمی‌آید که به ترتیب نمایانگر هجاهای بلند و هجاهای کوتاه هستند. بعد از این مرحله، رشته ورودی با استفاده از الگوریتم لونیشتاین با مجموعه‌ای از گروه‌های وزنی پرکاربرد زبان فارسی مقایسه می‌شود و رشته‌ای که کمترین فاصله یا به عبارت دیگر بیشترین مشابهت را با رشته ورودی داشته باشد به عنوان وزن عروضی انتخاب می‌شود. لازم به ذکر است که در نسخه اولیه برنامه، مواردی همچون «اختیارات شاعری»، «اختیارات وزنی»، و «عروض نیمایی» پشتیبانی نمی‌شود.

## ۲ تبدیل متن فارسی به زنجیره واجی

در این تحقیق از تحلیل‌گر صرفی پارس مورف (مواجی و دیگران، ۱۳۹۰) برای تبدیل متن فارسی به زنجیره واجی استفاده شده است. در طراحی پارس مورف، ساخت تصریفی کلمه در زبان فارسی (اسلامی و علیزاده‌لمجیری، ۱۳۸۸) به صورت ساختمان‌درا نظر گرفته شده است و جایگاه طبقات مختلف وندهای تصریفی در ساختمان انواع کلمات مشخص شده است. پارس مورف در تحلیل تصریفی کلمه ابتدا ستاک را شناسایی می‌کند و متناسب با ساخت تصریفی پیش‌بینی شده برای آن نوع ستاک به دنبال انواع وندهای تصریفی در جایگاه‌های خاص، با در نظر گرفتن صورت‌های مختلف نوشتاری هرکدام از طبقات، می‌گردد.

اگر ستاک پیچیده باشد، پارس مورف با اعمال قواعد واژه‌سازی زبان فارسی، ستاک مورد نظر را از حیث مشتق یا مرکب بودن تجزیه و تحلیل و اجزای سازنده آن را با ذکر مقوله دستوری و نقش آنها مشخص می‌کند. به دلیل معتبر نبودن فاصله به عنوان رمز کلمه در متون فارسی، پارس مورف در تجزیه ستاک‌های پیچیده، ترکیب‌های بالقوه را نیز به عنوان گزینه‌های بعدی در اختیار می‌گذارد.

در برنامه تحلیل‌گر صرفی پارس مورف این امکان وجود دارد که متن فارسی به عنوان ورودی برنامه داده شده و نتیجه به صورت زنجیره واجی نشان داده شود. برای علائم واجی زبان فارسی از علائم موجود در (ثمره، ۱۳۷۸) استفاده کرده‌ایم. مثلاً واج‌های  $\text{?}$ ،  $\text{š}$ ،  $\text{â}$ ،  $\text{ĵ}$  و  $\text{č}$  و  $\text{ž}$  به ترتیب نمایانگر نویسه‌های /همزه/، /ش/، /آ/، /اج/، /اچ/ و /اژ/ می‌باشند.

خروجی این برنامه برای جمله «آن یکی نحوی به کشتی درنشست» در جدول ۱ آمده است. همانطور که مشاهده می‌گردد، دو صورت واجی برای این عبارت نشان داده شده است که در یکی کلمه «کشتی» به صورت /kašti/ و در دیگری به صورت /kešti/ واج‌نویسی شده است.

جدول ۱: زنجیره واجی عبارت «آن یکی نحوی به کشتی در نشست»

ʔân yeki nahvi beh kašti dar nešast

ʔân yeki nahvi beh kešti dar nešast

در جدول ۲، زنجیره واجی عبارت «مردم حضور دارند» آمده است. در این حالت هم دو نوع واج‌نویسی برای کلمه «مردم» آمده است: اولی دارای مقوله دستوری اسم است که به صورت /mardom/ نوشته شده و دیگری نتیجه تحلیل صرفی به صورت فعل «مردن» در حالت اول شخص مفرد است یعنی mord+am. از آنجا که تحلیل نحوی روی عبارت ورودی انجام نمی‌شود، این زنجیره واجی نیز از لحاظ برنامه معتبر است. ولی با افزودن اطلاعات نحوی به برنامه می‌توان دقت آن را بالاتر برد.

جدول ۲: زنجیره واجی عبارت «مردم حضور دارند»

mardom hozur dârand

mordam hozur dârand

مسئله بعدی، مسأله کسره اضافه است که باید تمهیدی برای آن اندیشیده شود. چون کسره اضافه در متن فارسی نمایش داده نمی‌شود، ولی در زنجیره واجی حضور دارد؛ لازم است اطلاعات نحوی به برنامه تبدیل متن به زنجیره واجی اضافه شود تا دقت برنامه افزایش یابد. برای مثال در جدول ۳ خروجی برنامه به ازای عبارت «کتاب من کو؟» آورده شده که در آن کسره اضافه را نمی‌بینیم.

جدول ۳: زنجیره واجی عبارت «کتاب من کو؟»

ketâb man ku

## ۳ تقطیع

هجاهای فارسی از نظر امتداد سه نوعند: هجای کوتاه، هجای بلند، هجای کشیده (سلطانی طارمی، ۱۳۹۰). هجای کوتاه دارای دو واج است (صامت + مصوت کوتاه) و با (U) نمایش داده می‌شود. هجای بلند دارای سه واج است (و یا یک صامت + یک مصوت بلند) که با علامت (-) نمایش داده می‌شود. هجای کشیده بیش از سه واج دارد و با علامت (U-) نمایش داده می‌شود و از نظر امتداد معادل یک هجای بلند و یک هجای کوتاه است. در تعیین وزن شعر فارسی اگر وزن یک مصراع (در شعر کلاسیک) را به دست آوریم، وزن شعر به دست آمده است. علت آن است که در زبان فارسی هم مثل بسیاری از زبان‌ها، واحد وزن شعر مصراع است. یعنی وقتی شاعر مصراع اول شعر را سرود وزن شعرش را انتخاب کرده است. در بقیه مصراع‌ها همان وزن باید تکرار شود.

## ۱.۳ تقطیع هجایی

منظور از تقطیع هجایی آن است که مصراع‌های مورد نظر را به هجاهای تشکیل‌دهنده آن تقسیم کنیم. برای این کار ابتدا مصراع‌های مورد نظر را به خط عروضی می‌نویسیم، سپس آن را به هجاهای تشکیل‌دهنده‌اش تقسیم می‌کنیم. مثال در جدول ۴:

جدول ۴: مبنای دل به جز در یار خود دل / امید از هر که داری جمله بگسل

مَ	بَن	دِی	دِل	بِ	جُز	دَر	یَا	رِ	خُد	دِل	أ	مِی	دَز	هَر	کِی	دَا	رِی	جُم	لِ	بِگ	سِل
-	-	-	-	U	-	-	-	U	-	-	U	-	-	-	U	-	-	-	U	-	-

### ۲.۳ تقطیع به ارکان

در تقطیع به ارکان، هجاهای مصراع‌ها را سه‌تا سه‌تا یا چهارتا چهارتا یا سه‌تا چهارتا یا چهارتا سه‌تا تقسیم می‌کنند. تقسیم‌بندی‌های دیگر در عروض امروز رایج نیست. علت آن است که بیشتر کلمات فارسی حداکثر چهار هجا دارند و کلماتی که بیش از چهار هجا دارند بسیار کم‌اند. همچنین تقسیم‌بندی چهارتا سه‌تا بسیار نامتداول است. مثال در جدول ۵:

جدول ۵: باور مکن که من دست از دامنم بدارم / شمشیر نگسلاند پیوند مهربانان

بَا	وَر	مَ	اَکُن	کِی	مَنْ	اَس	تَر	دَا	مَ	اَت	بِ	دَا	رَم	شَم	شِیر	نَگ	سَ	لَا	نَد	بِی	وَن	دِ	مِه	زَ	بَا	نَان
-	-	U	-	U	-	-	-	-	U	-	U	-	-	-	U	-	U	-	-	-	-	U	-	U	-	-

حال می‌توان به جای هجاهای کوتاه و بلند از (-) و (U) استفاده کنیم. البته چون در صرف و نحو عربی همه کلمات را با «ف، ع، ل» می‌سنجند، در عروض عربی و فارسی هم وزن هجاهای تقطیع شده هر مصراع را (که نمایانگر نظم وزن‌اند) از «ف، ع، ل» ساخته‌اند (سلطانی طارمی، ۱۳۹۰). مثلاً هم‌وزن (U-) فَعولن است. همچنین به جای (—U) از قالب هم‌وزنش «مفاعیلن» استفاده می‌کنند. مهمترین قالب‌ها یا ارکان عروضی را وحیدیان کامکار نوزده تا دانسته است (وحیدیان کامیار، ۱۳۶۵).

### ۴ فاصله لونشتاین<sup>۱</sup>

فاصله لونشتاین (Levenshtein (1966) به صورت حداقل میزان تغییرات مورد نیاز برای رشته S به رشته T تعریف می‌شود که عملیات مجاز در آن عبارتند از: درج، حذف یا جانشینی یک نویسه واحد. برای مثال فاصله لونشتاین دو رشته apple و orange با یک ماتریس ساده در جدول ۶ محاسبه شده است. فاصله این دو رشته برابر ۵ است که در قسمت پایین و سمت راست ماتریس نشان داده شده است. این فاصله بدین معناست که رشته apple را با درج o، جانشینی r به جای a، a به جای p، n به جای p و g به جای l به رشته orange تبدیل می‌شود. یک عملیات درج و چهار عملیات جانشینی داریم که مجموعاً برابر ۵ می‌شود.

جدول ۶: مثالی از فاصله لونشتاین

		A	P	P	L	E
	۰	۱	۲	۳	۴	۵
O	۱	۱	۲	۳	۴	۵
R	۲	۲	۲	۳	۴	۵
A	۳	۲	۳	۳	۴	۵
N	۴	۳	۳	۴	۴	۵
G	۵	۴	۴	۴	۵	۵
E	۶	۵	۵	۵	۵	۵

<sup>۱</sup>Levenshtein distance

مراحل ساخته شدن ماتریس شکل ۱ به شرح ذیل است:

۱. ماتریس  $d$  با  $M$  سطر و  $N$  ستون ساخته می‌شود.
  ۲. به سطر اول مقادیر از 0 تا  $M$  و به ستون اول مقادیر از 0 تا  $N$  اختصاص داده می‌شود.
  ۳. هر نویسه‌ای از  $S$  (از 1 تا  $M$ ) و هر نویسه‌ای از  $T$  (از 1 تا  $N$ ) را بررسی می‌کنیم.
  ۴. اگر  $S[i]$  با  $T[j]$  برابر بود، هزینه برابر 0 و در غیر این صورت برابر 1 است.
  ۵. مقدار عنصر  $d[i, j]$  ماتریس برابر کمینه مقادیر زیر است:
    - (آ) عنصر بالایی بعلاوه یک:  $d[i - 1, j] + 1$
    - (ب) عنصر سمت چپ بعلاوه یک:  $d[i, j - 1] + 1$
    - (ج) عنصر بالا و سمت چپ بعلاوه هزینه:  $d[i - 1, j - 1] + cost$
  ۶. بعد از این که تکرار مراحل ۳، ۴ و ۵ به پایان رسید، مقدار فاصله در عنصر  $d[N, M]$  قرار دارد.
- شبه‌کد فاصله لونشتاین در شبه‌کد ۱ آمده است.

شبه‌کد ۱: فاصله لونشتاین

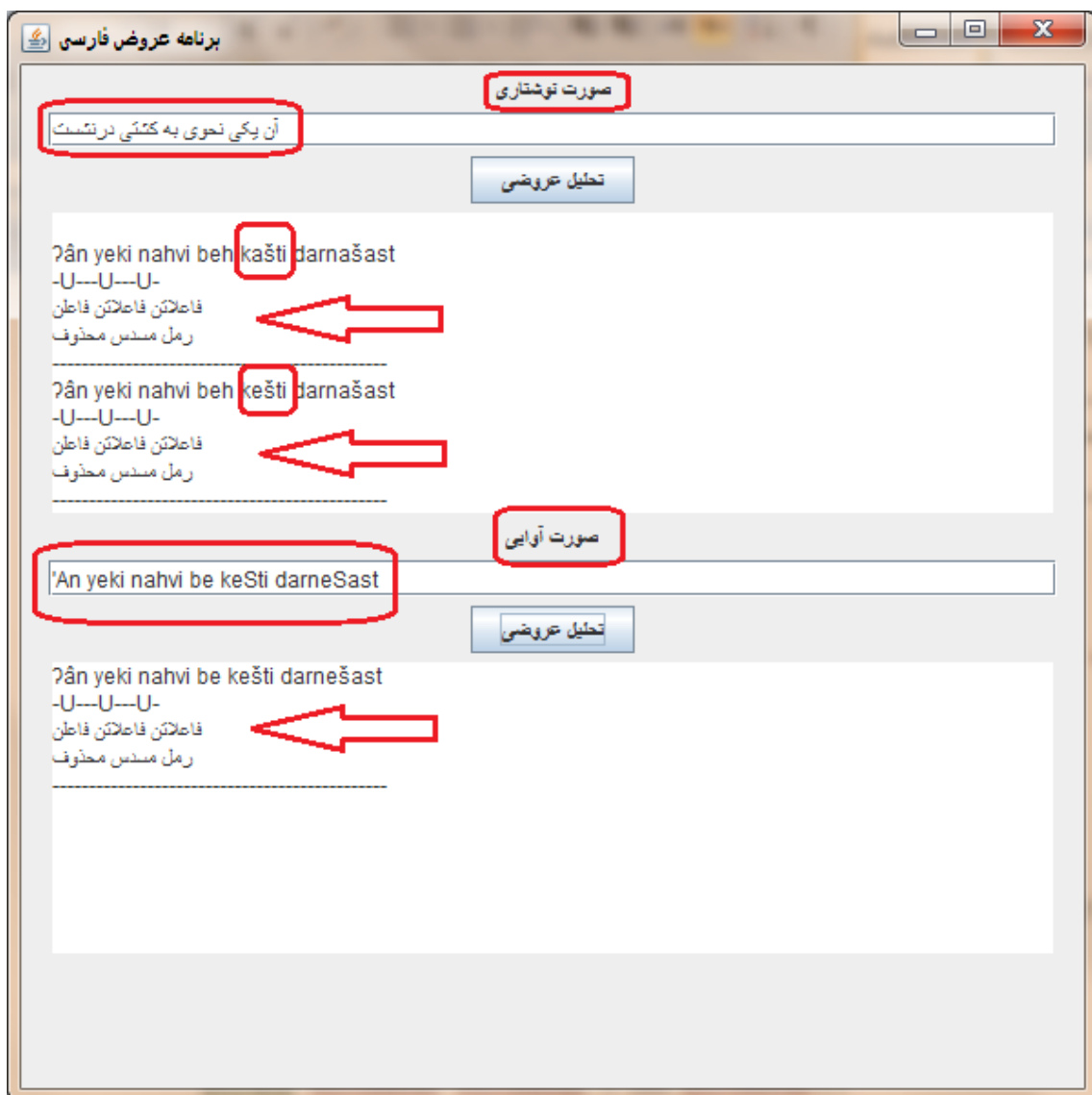
```
1 int LevenshteinDistance(char S[1..M], char T[1..N]){
2   declare int d[0..M, 0..N]
3   for i from 0 to M
4     d[i,0] := i //the distance of any first string to an
5                 empty second string
6   for j from 0 to N
7     d[0, j] := j //the distance of any second string to
8                 an empty first string
9   for j from 1 to N
10  {
11    for i from 1 to M
12    {
13      if S[i] = T[j] then
14        d[i, j] := d[i-1, j-1] //no operation
15                               required
16      else d[i, j] := minimum(d[i-1, j] + 1, //a
17                               deletion
18                               d[i, j-1] + 1, //an insertion
19                               d[i-1, j-1] + 1) //a substitution
20    }
21  }
22  return d[M, N]
23 }
```

## ۵ پیاده‌سازی

با توجه به مبانی نظری که به طور خلاصه در بخش‌های قبل ذکر شد، برنامه‌ای طراحی گردید که متن ورودی را گرفته و مراحل تقطیع هجایی و تقطیع به ارکان آن را انجام می‌دهد. البته اگر متن ورودی به خط فارسی باشد، ابتدا

به زنجیره واجی تبدیل می‌شود. با توجه به الگوی واجی زبان فارسی که به صورت‌های CV، CVC، CVCC است، هر گاه مصوتی پیدا شود، صامت قبل از آن متعلق به آن بوده و چون خوشه‌های همخوانی در آغاز هجا نداریم، لذا صامت‌های قبل از آن متعلق به هجای قبلی خواهند بود. بدین ترتیب مصراع ورودی به صورت رشته‌ای از (-)ها و (U)ها تبدیل می‌شود. سپس با استفاده از الگوریتم لونشتاین که در بخش ۴ شرح داده شد، در بین گروه‌های وزنی که در جدول ۷ آمده‌اند، آن گروهی که کمترین فاصله را با متن تقطیع شده دارد انتخاب کرده و به عنوان وزن شعر (مصراع) برمی‌گردانیم. در حقیقت، اگر هر کدام از گروه‌های وزنی را به صورت رشته‌ای از (-) و (U)ها در نظر بگیریم، آنگاه مسأله پیدا کردن وزن عروضی به مسأله پیدا کردن رشته‌ای با کمترین فاصله یا بیشترین مشابهت با رشته ورودی کاهش می‌یابد.

در شکل ۱ نمونه‌ای از خروجی برنامه را هم به صورت ورودی خط فارسی و هم ورودی واج‌نویسی شده برای رشته ورودی «آن یکی نحوی به کشتی درنشت» می‌بینیم که وزن «فاعلاتن فاعلاتن فاعلن» را به عنوان خروجی ارائه داده است. همانگونه که در شکل ۱ می‌بینیم، ابتدا متن ورودی به زنجیره واجی تبدیل شده، و بعد از اعمال تقطیع به صورت رشته‌ای از (-) و (U)ها درآمده است و نهایتاً با استفاده از الگوریتم لونشتاین، گروه وزنی آن مشخص شده است. به دلیل عدم نوشتن مصوت‌ها در خط فارسی و به تبع آن وجود نمودهای واجی مختلف برای یک رشته واحد ممکن است دقت الگوریتم بسته به نمود واجی که پیشنهاد می‌دهد تغییر کند یا کاهش یابد. به همین دلیل این امکان در برنامه فراهم شده است که شعر را به صورت زنجیره واجی نیز وارد نموده و وزن آن را تعیین کنیم.



شکل ۱: خروجی نمونه برنامه برای مصراع «آن یکی نحوی به کشتی درنشت»

جدول ۷: وزن‌های پرکاربرد در زبان فارسی

نام گروه وزنی	گروه وزنی	تقطیع
رمل مثنیٰ مخبون محذوف	فعلاتن فعلاتن فعلن	UU-UU-UU-UU-
مجتث مثنیٰ محذوف	مفاعیلن فعلاتن مفاعیلن فعلن	U-U-UU-U-U-UU-
مضارع مثنیٰ اخرب مکفوف محذوف	مفعول فاعلاتن مفاعیلن فاعیلن	-U-U-UU-U-U-
رمل مثنیٰ محذوف	فاعلاتن فاعلاتن فاعلاتن فاعیلن	-U-U-U-U-
هزج مثنیٰ سالم	مفاعیلن مفاعیلن مفاعیلن مفاعیلن	U-U-U-U-U
هزج مسدس محذوف	مفاعیلن مفاعیلن فعولن	U-U-U-U-
مضارع مثنیٰ اخرب	مفعول فاعلاتن مفعول فاعلاتن	-U-U-U-U-U-
هزج مثنیٰ اخرب مکفوف محذوف	مفعول مفاعیلن مفاعیلن فعولن	-UU-UU-UU-
خفیف مخبون محذوف	فعلاتن مفاعیلن فعلن	UU-U-U-UU-
رمل مسدس محذوف	فاعلاتن فاعلاتن فاعیلن	-U-U-U-U-
رجز مثنیٰ مطوی مخبون	مفتعلن مفاعیلن مفتعلن مفاعیلن	-UU-U-U-UU-U-U-
هزج مثنیٰ اخرب	مفعول مفاعیلن مفعول مفاعیلن	-UU-U-U-
رمل مثنیٰ مشکول	فعلات فاعلاتن فاعلاتن فاعلاتن	UU-U-U-UU-U-U-
مجتث مثنیٰ مخبون	مفاعیلن فعلاتن مفاعیلن فعلاتن	U-U-UU-U-U-UU-
هزج مسدس اخرب مقبوض محذوف	مفعول مفاعیلن فعولن	-UU-U-U-
مقارِب مثنیٰ اثلث	مستفعلاتن مستفعلاتن	-U-U-U-
منسرح مثنیٰ مطوی منحور	مفتعلن فاعلات مفتعلن فع	-UU-U-U-UU-
رجز مثنیٰ سالم	مستفعلن مستفعلن مستفعلن مستفعلن	-U-U-U-U-U-
منسرح مثنیٰ مطوی مکشوف	مفتعلن فاعیلن مفتعلن فاعیلن	-UU-U-UU-U-U-
مقتضب مثنیٰ مطوی مقطوع	فاعلات مفعولن فاعلات مفعولن	-U-U-U-U-U-
سریع مطوی مکشوف	مفتعلن مفتعلن فاعیلن	-UU-UU-U-
مقارِب مثنیٰ سالم	فعولن فعولن فعولن فعولن	U-U-U-U-U-
مقارِب مثنیٰ محذوف	فعولن فعولن فعولن فعل	U-U-U-U-U-
رجز مثنیٰ مطوی	مفتعلن مفتعلن مفتعلن مفتعلن	-UU-UU-UU-UU-
رمل مثنیٰ مخبون	فعلاتن فعلاتن فعلاتن فعلاتن	UU-UU-UU-UU-
رمل مسدس مخبون محذوف	فعلاتن فعلاتن فعلن	UU-UU-UU-
هزج مثنیٰ مکفوف محذوف	مفاعیلن مفاعیلن مفاعیلن فعولن	U-UU-UU-UU-
خفیف مثنیٰ مخبون	فعلاتن مفاعیلن فعلاتن مفاعیلن	UU-U-U-UU-U-U-U-
هزج مسدس اخرب مقبوض	مفعول مفاعیلن مفاعیلن	-UU-U-U-U-
رمل مثنیٰ سالم	فاعلاتن فاعلاتن فاعلاتن فاعلاتن	-U-U-U-U-U-
بسیط مثنیٰ مخبون	مستفعلن فعلن مستفعلن فعلن	-U-UU-U-UU-

## ۶ نتیجه‌گیری

در این مقاله، با استفاده از سامانه تحلیل‌گر صرفی کلمه در زبان فارسی با عنوان پارس مورف، که بر پایه یک مطالعه دقیق زبان‌شناختی از نظام صرفی زبان فارسی استوار است، سعی کردیم تا فرایند خودکار تبدیل متن فارسی به زنجیره واجی را توسعه دهیم و با استفاده از آن عمل تقطیع هجایی و تقطیع به ارکان را روی متن ورودی انجام دهیم. سپس با در نظر گرفتن تقطیع متن ورودی و گروه‌های وزنی پرکاربرد به صورت رشته‌ای از (-) و (U)ها و کاهش مسأله به مسأله یافتن کمترین فاصله بین دو رشته و بکارگیری الگوریتم لونشتاین، وزن عروضی مصراع



ورودی را به دست آوردیم.

از چالش‌های پیش رو که در کارهای آتی باید به آنها پرداخته شود، در نظر گرفتن دشواری‌های موجود در پردازش متن فارسی (اسلامی، ۱۳۸۱) است که باید راهکارهایی برای حل آن اندیشیده شود تا دقت الگوریتم تبدیل متن فارسی به زنجیره واجی و به تبع آن دقت مراحل تقطیع و یافتن وزن عروضی افزایش یابد. در کارهای آتی می‌توان مواردی همچون «اختیارات شاعری»، «اختیارات وزنی»، و «عروض نیمایی» را به امکانات برنامه افزود. از نتایج این تحقیق می‌توان به عنوان یک برنامه کمک آموزشی در زمینه‌های آموزش عروض به فارسی‌زبانان و غیر فارسی‌زبانان، بررسی صحت و دقت تعیین وزن عروضی اشعار به روش‌های سنتی و ارائه نحوه‌های مختلف خوانش اشعار کلاسیک فارسی اشاره نمود.

## مراجع

Levenshtein , V. I. (1966), "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," Soviet Physics Doklady, 10, 707. 4

Allen, J. (1992), "Overview of text-to-speech systems," in Advances in speech signal processing, eds. Furui, S. and Sondhi, M., New York: M. Dekker, bibtext: allen\_overview\_1992. 2

Torkkola, K. (1993), "An Efficient Way To Learn English Grapheme-To-Phoneme Rules Automatically," pp. 199 – 202 vol.2. 2

اسلامی، محرم، شریفی آتشگاه، مسعود، علیزاده لمجیری، صدیقه، و زندگی، طاهره (۱۳۸۳)، "واژگان زایای زبان فارسی"، در اولین کارگاه پژوهشی زبان فارسی و رایانه، دانشگاه تهران. ۱

اسلامی، محرم (۱۳۸۱)، "دشواری‌های پردازش رایانه‌ای خط فارسی"، نشر دانش، ۲۸ - ۳۲. ۹

اسلامی، محرم و علیزاده لمجیری، صدیقه (۱۳۸۸)، "ساختار تصریفی کلمه در زبان فارسی"، زبان و ادب فارسی، ۱۸-۱. ۲

ثمره، یدالله (۱۳۷۸)، آواشناسی زبان فارسی. مرکز نشر دانشگاهی، ویرایش دوم. ۲

سلطانی طارمی، سعید (۱۳۹۰)، عروض به زبان امروز. نشر نی. ۳، ۴

صیادیان، ابوالقاسم و نصیرزاده، مجید (۱۳۷۵)، "تجربه‌ای در مدل‌سازی زبان فارسی برای یک سیستم تبدیل متن به گفتار"، در دومین کنفرانس سالانه انجمن کامپیوتر ایران، صفحات ۱۰۵ - ۱۱۱. ۲

مواجی، وحید، اسلامی، محرم، و وزیرنژاد، بهرام (۱۳۹۰)، "پارس مورف: تحلیل‌گر صرفی زبان فارسی"، دوفصل‌نامه علمی-پژوهشی پردازش علائم و داده‌ها، ۸. ۲

وحیدیان کامیار، تقی (۱۳۶۵)، فنون و صنایع ادبی (عروض). نشر ایران. ۴