

فصلنامه تخصصی سبک‌شناسی نظم و نثر فارسی (بهار ادب)

علمی - پژوهشی

سال سوم - شماره چهارم - زمستان ۸۹ - شماره پیاپی ۱۰

ماده‌تاریخ و استفاده از رایانه در ساختن آن

(ص ۲۴۳ - ۲۳۱)

مهدی صادقی^۱، صفر وفادار دولق^۲ (نویسنده مسئول)

تاریخ دریافت مقاله: ۸۹/۹/۳

تاریخ پذیرش قطعی: ۸۹/۱۱/۶

چکیده:

با توجه به سابقه تاریخی شعر و ادب فارسی و علاقه‌مندی ایرانیان به ادبیات، لزوم ورود تحقیقات رایانه‌ای و علوم بین‌رشته‌ای در این زمینه اجتناب‌ناپذیر است. یکی از زمینه‌های شعر فارسی استفاده از ماده‌تاریخ در شعر است که تاکنون هیچ پژوهش رایانه‌ای در این زمینه انجام نشده است. البته کارهایی مانند پیدا کردن وزن عروضی یا موسیقی شعر انجام شده است. ماده‌تاریخ پدیده‌ایست که از قرن هفتم به بعد (البته به نظر برخی قرن پنجم) در ادبیات فارسی رواج یافته است که در آن تاریخهایی را در قطعه‌های کوتاه برای واقعه‌های گوناگون اعم از جلوس، فوت یا قتل پادشاهی و یا مرگ شاعر، حکیم و... ذکر میکنند. اصطلاحاً تاریخ‌گویی به این نحو را ماده‌تاریخ سازی می‌گویند. این مسئله شباهت بسیاری به مسئله کوله‌پشتی دارد. در این مقاله سعی شده است تا با استفاده از مفاهیم رایانه‌ای مسئله‌ای مربوط به ادبیات حل شود که نتایج قابل قبولی بدست آمده است.

کلمات کلیدی:

ماده تاریخ، حروف ابجد، تطبیق رشته، الگوریتمهای ابتکاری، داده کاوی، بسامد کلمات

۱ - استادیار، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری، تهران m-sadeghi@ibb.ut.ac.ir

۲ - دانشجوی کارشناسی ارشد، گروه الگوریتم و محاسبات، دانشگاه تهران، vafadar@ut.ac.ir

مقدمه:

امروزه دانش بسیاری در اسناد متنی موجودست که در داخل و خارج سازمانها ذخیره شده است. پایگاه‌داده‌های متنی، بدلیل افزایش اطلاعات در دسترس در قالب الکترونیکی، از قبیل کتابخانه‌های دیجیتال، انتشارات الکترونیکی و ...، بسرعت رشد میکنند. تکنیک ویژه‌ای از داده‌کاوی که کاویدن متن نامیده میشود، برای استخراج اطلاعات از مجموعه داده‌های متنی بکار میرود، گسترش یافته است. (Kantardzic : ۲۰۰۳). ما در این نوشتار سعی کرده‌ایم تا با استخراج اطلاعاتی که در دل آثار ادبی نهفته است، ترکیبات مناسب و زیبا برای ماده‌تاریخ بسازیم که در بخشهای بعدی توضیح داده خواهد شد. در گذشته حساب جمل در نزد منجمان و محاسبان و در تنظیم جداول نجومی و عددی معمول بود؛ ولی امروزه گذشته از علوم غریبه، استعمال آن تقریباً منحصر است به ساختن «ماده تاریخ» و شماره گذاری مقدمه و فهرست کتاب در نزد بعضی از مؤلفان (آیینه وند : ۱۳۷۷ : ۳۷۹). ماده تاریخ آنست که شاعر در شعر خود تاریخ واقعه‌ای مهم، سال ولادت و وفات یا موضوع مهم دیگری را با صراحت یا بصورت حساب جمل (ابجد) بیان کند. سعدی سال تالیف کتاب گلستان را چنین با صراحت بیان کرده است:

در آن مدت که ما را وقت خوش بود ز هجرت ششصد و پنجاه و شش بود
 مراد ما نصیحت بود و گفتیم حوالت بلا خدا کردیم و رفتیم
 ملک‌الشعرای بهار ماده‌تاریخ زیر را در مرگ ایرج میرزا گفته است:

سکته کرد و مرد ایرج میرزا قلب ما ففسرد ایرج میرزا
 گفت بهر سال تاریخش بهار «و چه راحت مرد ایرج میرزا»
 که عدد ابجد (حساب جمل) «و چه راحت مرد ایرج میرزا» عدد ۱۳۴۴ میشود که سال مرگ ایرج میرزا سال (۱۳۴۴ ه.ق) است.

همچنین استاد همایی در مرگ بهار چنین ماده تاریخی را بکار برده‌اند:

کلک مشکین شنا از بهر تاریخش نوشت «گلشن علم و ادب بفسرد با مرگ بهار»
 عدد ابجد «گلشن علم و ادب بفسرد با مرگ بهار» ۱۳۷۰ یعنی سال مرگ بهار میباشد. (اسفندیارپور : ۱۳۸۸ : ۲۸۳)

در لغتنامه دهخدا در ذیل کلمه ماده تاریخ آمده است: ماده تاریخ کلمه یا کلماتی که به حساب جمل مساوی تاریخ مطلوب برآید و آن کلمه یا کلمات با معنی باشد. [در واقع باید] مجموع حروف بیت یا مصراع یا عبارتی به حساب ابجد با تاریخ واقعه‌ای تطبیق کند. (دهخدا: روایت دوم)

گروهی از طرفداران ادبیات تمایل چندانی به استفاده از صنایع بدیعی ندارند و آن را جدا از روح لطیف شعر میدانند و از پیچیدگی و بی‌فایده بودن آن سخن به میان می‌آورند. البته شکی در آن نیست که اینگونه صنایع ادبی هرگز جایگاه مفاهیم و مضامین ادبی را نخواهند گرفت اما همواره آدمی در پی تنوع و زیباییهای مختلف و پیچیدگی و رازگونه بودن می‌گردد. در واقع شعر نیز محصول این حس درونی آدمیست که جملات ساده را بصورت پیچیده و کلمات پس و پیش شده بیان کند و در عوض به جمله آهنگ، شگردها و هنرمندیهای لفظی و معنوی ببخشد. شاعران بزرگی که بر قله‌های ادب فارسی ایستاده‌اند نیز از این قاعده مستثنی نبوده‌اند و از این صنایع بهره جسته‌اند. ما در این نوشتار سعی خواهیم کرد تا به یکی از صنایع ادبی، یعنی ماده تاریخ بپردازیم. از آنجایی که ساختن ماده تاریخ، کاری بسیار دشوارست، ما سعی کرده‌ایم با استفاده از رایانه و استفاده از الگوریتمهای مناسب، علاوه بر ساختن ماده تاریخ، نگرش تازه‌ای به برخی مفاهیم ادبی با دیدی رایانه‌ای داشته باشیم. مفاهیمی از قبیل پیچیدگی محاسباتی، تطبیق (تراز) کلمات، بسامد تکرار کلمات و بسامد همنشینی کلمات را مورد بررسی قرار خواهیم داد.

۲- حروف ابجد

صور هشتگانه حروف جمل با نام اولین صورت (ابجد) معروفست. ترتیب حروف (صامت: بی صدا) در این نسق همان ترتیب عبری- آرامی است و این امر مؤید آنست که عرب، الفبای خود را از آنان (به وساطت قبطیان) اقتباس کرده است. (معین: ۱۳۷۵). حروف ابجد به این شرحند: «الف، ب، ج، د، ه، و، ز، ح، ط، ی، ک، ل، م، ن، س، ع، ف، ص، ق، ر، ش، ت، ث، خ، ذ، ض، ظ، غ». و اعداد ۱ تا ۱۰۰۰ به بیست و هشت حرف تقسیم شده است. برای آسانی از بر کردن ترتیب این حروف، هر چندتا حرف پیاپی به شکل واژه در آمده و تلفظ میشوند. این واژه‌ها عبارتند از: «أَبْجَدُ - هَوَزُ - حُطَى - كَلَمَنُ - سَعْفَصُ -

قَرَشَتْ - تَخَذُ - ضَطَّغُ. (مصاحب : ۱۳۷۸ : ۵) برای هر کلمه که بخواهند به عدد تبدیل کنند، عدد متناظر را از جدول (۱) برمی‌دارند و اعداد بدست آمده را جمع می‌زنند. برای نمونه در این روش، واژه «علی» برابر با «۱۱۰» میشود همچنین «یا علی» برابر با ۱۲۱ میباشد.

جدول (۱) عدد ابجد مربوط به هر حرف

الف	ب	ج	د	ه	و	ز	ح	ط	ی	ک	ل	م	ن	س	ع	ف
۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۲۰	۳۰	۴۰	۵۰	۶۰	۷۰	۸۰
ص	ق	ر	ش	ت	ث	خ	ذ	ض	ظ	غ						
۹۰	۱۰۰	۲۰۰	۳۰۰	۴۰۰	۵۰۰	۶۰۰	۷۰۰	۸۰۰	۹۰۰	۱۰۰۰						

همزه (أ) و الف (ا) هر یک را به عدد (۱) حساب میکنند و حروفی که بواسطه تشدید مکررست، یک حرف محاسبه میشوند؛ مثلاً کلمه فَعَال را معادل (۱۸۱) میگیرند؛ چون (ف معادل ۸۰)، (ع معادل ۷۰)، (ا معادل ۱) و (ل معادل ۳۰) است. استثنائاً لام مشدد کلمه الله را دو حرف حساب میکنند. بنابراین الله (۶۶) میشود. حروف ابجد کاربردهای دیگری نیز دارد. بعنوان مثال «در موسیقی پس از اسلام، در ثبت اصوات و دساتین از حروف ابجد استفاده شده است. برای نمونه نغمه‌های دستگاه شور با حروف ابجد عبارت بوده‌اند از: «یح یه یب ی ح ه ج ا» و فاصله‌ها و ابعاد آنها: «ط ط ج ج ط ج ج» که در موسیقی امروز به ترتیب نت‌های سل، لاکرن، سی بمل، دو، رکن، می بمل، فا و سل را تشکیل میدهند.» (ستایشگر : ۱۳۸۱ : ۳۷۵) برخی افراد برای این شماره‌ها ویژگی‌های فراطبیعی قائلند و تفسیرهای مختلفی درباره این اعداد ارائه میدهند که ما در این نوشتار تنها به جنبه ادبی و کاربرد آن در ادبیات خواهیم پرداخت و آن را تنها بعنوان یک ابزار در ادبیات بررسی خواهیم کرد.

۳- ماده تاریخ

یکی از ترفندهای شاعرانه که آن را جزء آرایه‌های ادبی آورده‌اند، ماده‌تاریخ است. دکتر رزمجو در کتاب انواع ادبی و آثار آن در زبان فارسی تعریف ماده تاریخ و نمونه‌هایی را

ذکر کرده‌اند. برای نمونه ماده تاریخ جلوس محمدعلی شاه قاجار که ادیب الممالک فراهانی در قطعه طنزآمیز ذیل آن را چنین به نظم درآورده‌اند:

صاحب تاج گشت و غاصب تخت تا محمدعلی شاه قاجار
 سیل کین کند از عمارت داد پایه استوار و ریشه سسخت
 نه سار ماند در جهان نه وثاق نه گیا هشت بر زمین نه درخت
 بس که بدبخت بود اهل هنر همه بستند از پناهاش رخت
 گشت سال جلوس او به سریر بی کم و کاست «ای شه بدبخت»

که عبارت «ای شه بدبخت» به حساب ابجد برابرست با ۱۳۲۴ (رزمجو: ۱۳۷۴: ۱۴۸). همچنین نخجوانی در کتاب مواد التواریخ خود، ماده تاریخ را چنین تعریف میکند: فن ماده- تاریخ از ظرائف فنون شعریه است و اغلب آن را جزو صنایع مستظرفه مانند موسیقی و نقاشی و امثال آنها دانسته‌اند و مربوطست به حساب ابجد یا حساب جمل، و در زمانهای بسیار قدیم نزدیک به اوایل اسلام به حساب ابجد یعنی با حروف مقطعه تاریخ گفتن معمول بود. بعداً حساب جمل در میان شعرا و ادبای عرب و عجم جاری و معمول گردیده و عمده شیوع آن از قرن پنجم هجری بوده است. (نخجوانی: ۱۳۴۳:) بعنوان مثال در مورد تاریخ وفات «رود»، حافظ در قطعات خویش (قطعه ۳۱) چنین سروده است:

آن میوه بهشتی کامد بدستت ای جان در دل چرا نکشتی، از دست چون بهشتی
 تاریخ این حکایت گر از تو باز پرسند سر جمله اش فرو خوان از «میوه بهشتی»
 جمع ابجدی دو کلمه «میوه بهشتی» معادل عدد (۷۷۸) است و نشان میدهد حادثه مرگ «رود» فرزند جوان حافظ در سال ۷۷۸ هجری قمری بوده است یعنی در سن (۵۲) سالگی حافظ.

۴- الگوریتم یافتن کلمه برای عدد

همانطور که از مثالهای مطرح شده مشخص است، ماده تاریخ بدین گونه ساخته میشود که شاعر عددی (معمولاً تاریخ) را در نظر گرفته و سعی میکند کلمه یا کلماتی را با استفاده از حروف بگونه‌ای شکل دهد که حساب ابجدی آنها، عدد مورد نظر را ایجاد کند. پیدا کردن عدد برای کلمه یا کلمات داده شده بسیار ساده بوده و فقط کافی است اعداد مربوط به هر حرف را جمع بزنیم. اما ساختن کلمه یا عبارتی که کد ابجد آن عددی مانند (n) شود، کار

ساده‌ای نیست. تعداد حالات ممکن بسیار زیادست بررسی کردن معنادار یا بی‌معنی بودن آنها در زمان چند جمله‌ای قابل حل نمیباشد. این مسئله شباهت بسیاری به مسئله کوله‌پشتی که (NP-) کاملست، دارد اما این مساله ذاتا نمایی است. بعنوان مثال عدد (۸۹) را از ترکیب «ف»+«ط» (یعنی ۹+۸۰) یا «ف»+«ح»+«ا» یا «ف»+«ز»+«ب» یا «ه»+«م»+«د»+«م» یا ... ساخت. همانطور که مشاهده میشود تعداد حالات ممکن بسیار زیاد و بسیاری از آنها بی‌معنی است. اما برخی از آنها مانند «ه»+«م»+«د»+«م» میتواند کلمه «همدم» را بسازد. البته جایگشت‌های حروف تغییری در کد ایجاد آنها نمیکند اما معنی را تغییر میدهد. از این دسته میتوان به «امید دل»، «جلوه مه» و «آینه پاک» اشاره کرد. با توجه به توضیحات داده شده میتوان نتیجه‌گیری کرد که این مسئله از مسئله کوله‌پشتی سخت‌تر است و چون مسئله کوله‌پشتی (NP-) کاملست پس مسئله یافتن کلمه برای عدد دلخواه مطابق با حساب جمل (کد ایجاد) نیز در زمان چند جمله‌ای قابل حل نیست. البته در برخورد با مسئله یافتن کلمه برای عدد، با در نظر گرفتن نکاتی میتوان جوابهای قابل قبولی را در زمان مناسب بدست آورد. در ادامه با توضیح برخی مفاهیم، سعی خواهیم کرد راه حلی برای این مساله ارائه دهیم.

۴-۱- دامنه کلمات

طبق تعریف دهخدا، کلمه یک جزو از کلام، لفظ معنی‌دار [است] و فرق کلمه با لفظ در اینست که لفظ اعم است از معنی‌دار و بی‌معنی ولی کلمه حتما معنی دارد. (دهخدا) تعداد حروف زبان محدودست ولی از این حروف محدود میتوان کلمات بیشماری ساخت. اما در هر زبانی تعداد کلمات زنده‌ای که بکار برده میشود عملا محدودست و همچنین هر فردی دامنه کلمات محدودی دارد و در بین آن کلمات برخی را بیشتر و برخی را کمتر بکار میبرد. بعنوان مثال لغتنامه دهخدا (روایت دوم- انتشارات دانشگاه تهران) (۳۴۳۴۶۶) مدخل دارد که برخی از آنها جزو کلمات محسوب نمیشوند. یا بعنوان مثال دیگر، تعداد کل کلمات مختلف بکار رفته در دیوان حافظ کمتر از نه هزار کلمه میباشد. (این اعداد با استفاده از برنامه‌ای که برای شمارش کلمات مختلف در یک متن، نوشته شد بدست آمد که در آن برخی ترکیبات که معنا را تا حدی تغییر میدهند نیز بعنوان یک کلمه جدید در نظر گرفته

شده است. اگر ترکیبات در نظر گرفته نشوند تعداد کلمات مختلف از هفت هزار کلمه نیز کمتر خواهد شد و اگر افعال و صفات بدست آمده از بن‌های افعال را نیز کم کنیم تعداد کلمات بشدت کاهش خواهد یافت.

۴-۲- بسامد کلمات

کلمات در هر زبان و برای هر فردی بسامدهای مختلفی دارند. برخی کلمات پرتکرار و برخی دیگر کم‌تکرار هستند. بعنوان مثال در جدول (۲) بسامد تکرار برخی کلمات در دیوان حافظ، نمایش داده شده است. از این بسامدها برای مرتب‌سازی و ساختن کلمات ترکیبی استفاده شده است.

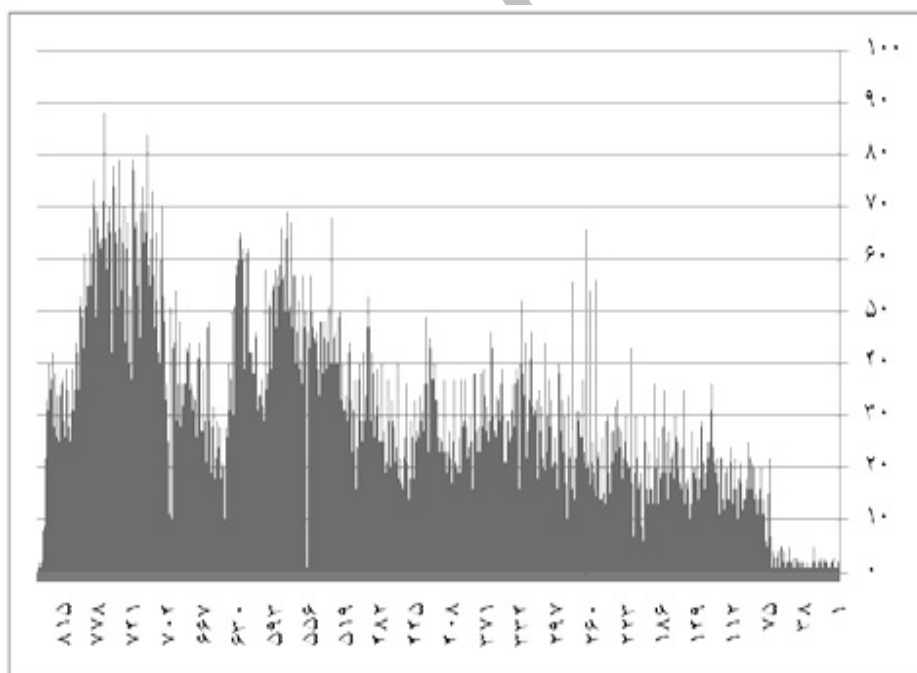
جدول (۲) بسامد بعضی کلمات در دیوان حافظ

کلمه	بسامد
حافظ	۴۸۷
حضور	۱۴
یار	۱۹۵
مریم	۱
سمن	۱۳
یاسمین	۱
ارغوان	۹
ارغوانی	۳
ارغنون	۲
عقل	۴۰

۴-۳- الگوریتم

با توجه به توضیحات بالا، میتوان تمهیداتی در نظر گرفت تا بتوان جوابهای قابل قبولی پیدا کرد. بعنوان مثال نیازی نیست که تمام ترکیبات ممکن حروف برای ایجاد کلمات در نظر گرفته شود. کافیست با در نظر گرفتن یک پایگاه داده از کلمات و محاسبه کد ابجد برای هر کلمه، یکبار محاسبات مربوط به کد ابجد را انجام داد و در هنگام جستجوی کلمه یا

کلماتی برای عددی خاص، طبق اختیاراتی که برای کاربر در نظر گرفته میشود، پردازش مربوطه انجام پذیرد. تعداد پاسخهای یافته شده به تعداد کلمات داخل پایگاه داده و نحوه توزیع آنها (بر اساس کد ابجد) بستگی دارد. برای پر کردن پایگاه داده‌ها، از مجموعه لغات **Lexicon** استفاده شده است. این مجموعه از یک برنامه متن باز به نام «واژه‌های زبانی زبان فارسی» برداشته شده است. علاوه بر آن از برخی کلمات و ترکیبات دیوان اشعار شعرایی مانند: حافظ، خیام، مولانا و برخی شعرای معاصر مانند بهار، رهی معیری و ایرج میرزا نیز استفاده شده است. نمودار توزیع کلمات پایگاه داده بر اساس کد ابجد در شکل (۱) نمایش داده شده است. بعنوان مثال پایگاه داده ساخته شده، دارای شصت و چهار کلمه (کلماتی از قبیل: «اندرز»، «برهنه»، «ایران»، «جبارانه»، «جابرانه» و ...) با کد ابجد (۲۶۲) میباشد.



شکل (۱): نمودار توزیع فراوانی کلمات پایگاه داده بر اساس کد ابجد

در زیر دستور (SQL) مربوط به یافتن کلمه یا کلمات برای عددی خاص نمایش داده شده است. البته با توجه به بزرگی جداول و زمان‌بر بودن ضرب دکارتی در (SQL)، از دستور (SQL) استفاده نکرده‌ایم، و سعی کرده‌ایم با بهینه کردن برخی دستورات و

گذاشتن برخی شرطها، سرعت جستجو و میزان حافظه مصرفی را کاهش دهیم. ماهیت الگوریتم تغییر نکرده است و ما برای سادگی، همان دستور (SQL) را نمایش میدهیم. (abjadGoal) کد ابجد مورد نظر، (W_1, W_2, \dots, W_w) جداول کلمات هستند. حداکثر تعداد کلمات یعنی w عددیست که از کاربر سؤال میشود و با توجه به آن جداول (W_1, W_2, \dots, W_w) ایجاد میشوند. بعنوان مثال کاربر بدنبال عدد (۱۳۸۹) میگردد و میخواهد حداکثر دو کلمه‌ای باشد. کد ابجد عبارتهای «فکر شاعران جهانپرو»، «شاعران درست پیمان» و «اندیشمند آخرالزمان» عدد (۱۳۸۹) میشود ولی چون کاربر حداکثر عبارات دو کلمه‌ای را میخواهد، عبارت «فکر شاعران جهان پرو» ساخته نخواهد شد.

```
SELECT *
FROM W1, W2, ..., Ww
WHERE (W1.abjadCode + W2.abjadCode + ... + Ww.abjadCode) = abjadGoal
```

برای پایگاه داده‌ای با ده هزار کلمه نیز جوابهای بسیاری یافت میشود. بعنوان مثال ممکن است برای عددی مانند (۲۶۲) با ترکیبات حداکثر دو کلمه‌ای، بیش از شش هزار مورد یافت شود. (به این دلیل عدد دقیق گفته نمیشود چون به کلمات داخل پایگاه داده بستگی دارد). آرایه بیش از ۶۰۰۰ کلمه یا ترکیب برای کسی که قصد ساختن ماده تاریخ دارد، بسیار مفید خواهد بود اما گشتن در میان ۶۰۰۰ کلمه شاید چندان دلپذیر نباشد و کاربر تمایل داشته باشد تا نتیجه این جستجو مرتب شود. در واقع ما باید برای نتایج یافته شده معیارهایی جهت مرتب سازی داشته باشیم. معیارهای مختلفی را میتوان برای نتایج بدست آمده، در نظر گرفت. معیارهایی از قبیل طول رشته، تعداد کلمات تشکیل دهنده، و همچنین آرایه‌های ادبی مانند واج آرای، جناس و ... که در ادامه خواهد آمد.

۴-۵- واج آرای

واج آرای تکرار یک یا چند واج صامت یا مصوت در شعر یا در نثرست در کلمه‌های یک مصراع یا بیت بگونه‌ای که آفریننده موسیقی درونی باشد و بر تأثیر شعر بیفزاید. واج آرای یکی مفهوم کیفی است و برای مقداردهی به آن باید تا حدی از پیچیدگی تعریف آن کاست.

برای هر رشته عددی نسبت داده شده است که میتواند بیانگر واج‌آرایی، در آن رشته باشد. در این برنامه، از پارامتر ساده‌ای استفاده شده است. نسبت طول هر رشته، به حروف مختلف بکار رفته شده در آن رشته (که عددی بزرگتر یا مساوی یک است) معیاری برای اندازه‌گیری واج‌آرایی در نظر گرفته شده است. بعنوان مثال برای رشته «مرغان هوایی» مقدار $\frac{10}{9}=1.11$ (این رشته به طول ۱۰ و ۹ دارای حرف مختلف است. حرف «ا» دوبار تکرار شده است.) و برای رشته «شور شراب عشق» مقدار $\frac{10}{7}=1.43$ خواهد بود. البته این پارامتر بسیار ساده و تا حدی وابسته به طول رشته است. برای سنجش بهتر میتوان از رابطه‌های پیچیده‌تری استفاده کرد. بعنوان مثال میتوان از تقسیم مجموع مربعات تعداد تکرار حروف بر طول رشته استفاده کرد. طبق رابطه (۱) برای رشته «مرغان هوایی» مقدار $\frac{1}{4}$ و طبق رابطه (۲) برای رشته «شور شراب عشق» مقدار $\frac{1}{8}$ بدست خواهد آمد.

$$\frac{1^2+1^2+1^2+2^2+1^2+1^2+1^2+2^2}{10}=\frac{14}{10}=1.4 \quad (1)$$

$$\frac{3^2+2^2+1^2+1^2+1^2+1^2+1^2}{10}=\frac{18}{10}=1.8 \quad (2)$$

۶-۴- جناس

جناس یا تجنیس، از صنایع لفظی در علم بدیعت است. تجنیس در لغت بمعنای «همجنس آوردن» و «با چیزی مانند شدن» است و در اصطلاح همانندی دو کلمه در لفظ با اختلاف آنها در معناست، مانند «شیر» که هم به شیر آشامیدنی و هم به جانور درنده گفته میشود. برای پیدا کردن جناس، میتوان از کاری که در زیست‌شناسی برای تراز کردن رشته‌های (DNA) استفاده میشود، استفاده کرد. الگوریتمهای تطبیق رشته‌ای، دسته مهمی از الگوریتمهای رشته‌ای هستند که سعی میکنند محل رخداد یک یا چند رشته (الگو) را در یک رشته بزرگتر پیدا کنند. با استفاده از تطبیق رشته‌ها میتوان آرایه جناس را در یک عبارت تشخیص داد. بعنوان مثال بیت زیر از بوستان سعدی را در نظر بگیرید:

یکی شاهی در سمرقند داشت که گفتی به جای سمرقند داشت

بعد از تطبیق دو مصرع میتوان شباهت دو رشته را اندازه گرفت. نتیجه این تطبیق در جدول (۳) نشان داده شده است.

جدول (۳) : تطبیق دو رشته

ی	ک	ی	ش	ا	ه	د	ی	د	ر	x	x	x	س	م	ر	ق	ن	د	د	ا	ش	ت
x	ک	x	ه	گ	ف	ت	ی	ب	ه	ج	ا	ی	س	م	ر	ق	ن	د	د	ا	ش	ت

۴-۷- مرتب سازی نتیجه

با توجه به تعاریف ارائه شده، به هر یک از رشته‌ها (کلمه یا کلمات یافته شده)، عددی بعنوان امتیاز اختصاص داده میشود و در خاتمه با توجه به امتیازهای اختصاص داده شده، نتایج مرتب شده و نمایش داده میشود. در رابطه (۳) پارامتر $score$ امتیاز هر رشته، W_{PUN} وزن امتیاز جناس و $PUNscore$ امتیاز جناس هر رشته، $W_{PHONEME}$ وزن امتیاز جناس و $PHONEMEscore$ امتیاز واج آرایشی آن رشته است. $POETscore$ امتیاز رشته را نسبت به بسامد کلمات آن رشته در دیوان شاعران و با توجه به ارزش دهی کاربر در قسمت تنظیمات برنامه و در بخش میزان تاثیر شاعران که در رابطه (۴) آمده را نشان میدهد. $FREQUENCYscore$ امتیاز رشته را با توجه به بسامد کلمات آن رشته و با توجه به اطلاعات پایگاه داده مجموعه لغات $Lexicon$ محاسبه میکند. W_{POET} وزن امتیاز بسامد تکرار کلمات توسط شاعران و $W_{FREQUENCY}$ وزن امتیاز بسامد رشته را نشان میدهد.

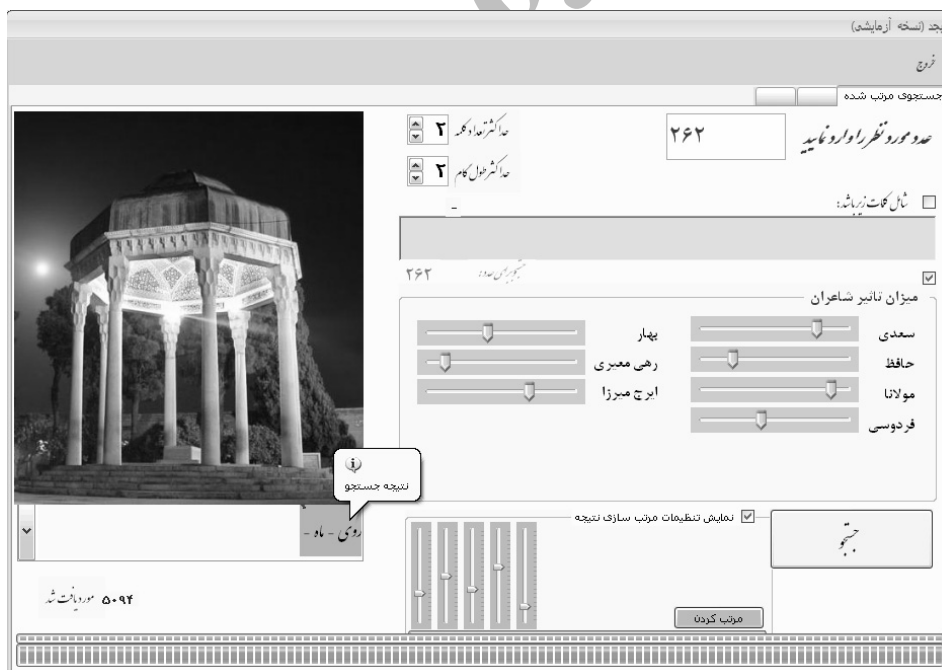
$$score = W_{PUN} \times PUNscore + W_{PHONEME} \times PHONEMEscore + W_{POET} \times POETscore + W_{FREQUENCY} \times FREQUENCYscore \quad (3)$$

$$POETscore = W_{hafez} \times F_{hafez} + W_{sadi} \times F_{sadi} + W_{molana} \times F_{molana} + W_{ferdosi} \times F_{ferdosi} + W_{khayyam} \times F_{khayyam} + W_{araqi} \times F_{araqi} + W_{nima} \times F_{nima} + W_{sohrab} \times F_{sohrab} + W_{shamlu} \times F_{shamlu} + W_{forugh} \times F_{forugh} + W_{akhavan} \times F_{akhavan} + W_{ebtehaj} \times F_{ebtehaj} \quad (4)$$

در رابطه (۴)، (W_x) نشان دهنده وزن امتیاز بسامد کلمات رشته برای شاعر (X) میباشد. این وزن با توجه به تنظیمات داخل برنامه، عددی از (۰ تا ۲۰) است که توسط کاربر برای هر شاعر تعیین میشود. (F_x) نشان دهنده مجموع بسامد کلمات رشته با توجه به آثار در نظر گرفته شده برای شاعر (X) میباشد.

۷- پیاده‌سازی

برنامه^۱ با (C#) در محیط (Net Farmework ۳.۵) پیاده‌سازی شده است. برنامه پیاده‌سازی شده در شکل (۳) نمایش داده شده است. در قسمت بالایی سمت راست برنامه، کاربر عدد مورد نظر خود را وارد و پس از انجام تنظیمات دلخواه خود، دکمه جستجو را می‌زند. پس از انجام محاسبات و جستجو، نتیجه در سمت چپ برنامه نمایش داده میشود. تنظیمات مربوط به میزان تاثیر شاعران، فقط ترتیب نمایش نتیجه را تغییر میدهد.



شکل (۲) برنامه پیاده‌سازی شده

۸- نتیجه :

استفاده از تمامی مسایل مطرح شده، نمیتواند جایگزین انسان در ادبیات شود، اما میتوان آن را بعنوان ابزاری در اختیار شاعر قرار داد تا بتواند با استفاده از آن، در زمان کوتاهی، به آنچه که در ذهن خویش دارد، برسد. روشهای رایج شده در این نوشتار، میتواند به شاعر کمک کند تا برای شعر گفتن، دامنه وسیعتری از کلمات را در اختیار داشته باشد. با توجه به کارهای رایانه‌ای انجام شده قبلی در ادبیات مانند [۷] و [۸]، امیدواریم استفاده از رایانه در حوزه ادبیات باعث پیشرفت و سهولت در امر تحقیق و مطالعه در این حوزه شود.

فهرست منابع :

- ۱- آیین‌وند، صادق (۱۳۷۷)، علم تاریخ در گستره تمدن اسلامی، ج ۱، تهران: پژوهشگاه علوم انسانی.
- ۲- اسفندیار پور، هوشمند (۱۳۸۸)، عروسان سخن، تهران: نشر فردوس.
- ۳- مصاحب، غلامحسین (۱۳۷۸)، و دیگران، دایرةالمعارف فارسی، تهران: نشر شرکت انتشارات علمی و فرهنگی.
- ۴- معین، محمد (۱۳۷۵)، فرهنگ فارسی، تهران: انتشارات امیر کبیر.
- ۵- موسسه لغت‌نامه دهخدا (بی تا)، لغت‌نامه دهخدا، تهران: نشر موسسه انتشارات و چاپ دانشگاه تهران، روایت دوم.
- ۶- نخجویی، حسین (۱۳۴۳)، مواد التواریخ، تهران: نشر ادبیه .
- ۷- عظیمی، محمدجواد، مقایسه موسیقی شعر حافظ، خواجه و سلمان با استفاده از کامپیوتر، دومین کارگاه پژوهشی زبان فارسی و رایانه، ص ۲۲۶، ج ششم و هفتم تیر ۱۳۸۵ .
- ۸- مجیری، محمد مهدی، سیستم هوشمند تشخیص وزن عروضی اشعار فارسی، مقاله کارشناسی، دانشگاه کاشان، ص ۹۲، ۱۳۸۷
- ۹- رزمجو، حسین (۱۳۷۴)، انواع ادبی و آثار آن در زبان فارسی، چاپ سوم، انتشارات آستان قدس رضوی.
- ۱۰- ستایشگر، مهدی (۱۳۸۱)، واژه‌نامه موسیقی ایران زمین، جلد اول. چاپ دوم، تهران: اطلاعات.
- ۱۱-

Kantardzic M, Data Mining: Concepts, Models, Methods, and Algorithms, John Wiley & Sons, ۲۰۰۳