# روش جدید خلاصه سازی استخراجی تک سندی با استفاده از نظریه مرکزیت

حسین کامیار، محسن کاهانی، محسن کامیار، آصف پورمعصومی حسن کیاده

**Hossein-kamyar@stu-mail.um.ac.ir, kahani@um.ac.ir, {mo_ka13, as.pormadoomi}@stu-mail.um.ac.ir**

آزمایشگاه فناوری وب، دانشگاه فردوسی مشهد

## چکیده

**در این مقاله یک تکنیک جدید برای خلاصه کردن یک متن با توجه به ویژگی‌های زبانی عناصر موجود در متن و زنجیره‌های معنای[1] میان آن‌ها ارائه شده‌است. در بسیاری از روش‌های موجود خلاصه‌سازی متون، تمامی توجه به ویژگی‌های آماری عناصر متن می‌باشد. در اینجا با استفاده از نظریه مرکزیت[2] که زنجیره‌های انسجام در داخل متن را تشخیص می‌دهد، یک روش جدید خلاصه‌سازی خودکار متن ارائه می‌شود. برای پردازش متن توسط نظریه مرکزیت و استخراج یک خلاصه‌ی منسجم احتیاج به در-هم آمیختن تعدادی از ابزارهای پردازش متن شامل روش جایگزینی ضمایر و گروه‌های اسمی با مرجع اسمی آنها (Coreference Resolution)، روش برچسب‌زنی نقش گرامری اسامی در داخل یک جمله (Semantic Role Labeling) و روش برچسب‌زنی گروه‌های اسمی (POS Tagger)[3] می‌باشد.**

**در این مقاله، نتایج تجربی به دست‌آمده از اعمال نظریه مرکزیت برای خلاصه‌سازی متون بر روی یک پیکره‌ی خلاصه‌سازی ارائه می‌گردد. با ارزیابی روش‌های خلاصه‌سازی موجود و الگوریتم پیشنهادی بر روی این پیکره، می‌توان به میزان افزایش کارآیی و کیفیت حاصل‌شده، پی برد.**

**کلمات کلیدی:** خلاصه‌سازی تک‌سندی، نظریه‌ی مرکزیت، Coreference Resolution، SRL

## ۱- مقدمه

خلاصه‌سازی خودکار متون یکی از ابزارهای مهم و در عصر افزایش انفجارگونه ی اطلاعات می‌باشد. طبق تعریف [1] خلاصه به یک متن تولیدشده از یک یا چند متن دیگر اطلاق می‌شود که دربردارنده مفاهیم مهم متن یا متون مبدأ می‌باشد. این متن تولیدشده نباید از نصف متن یا متون مبدأ بزرگتر باشد. همین تفسیر ساده دربردارنده ویژگی‌های اصلی یک خلاصه می‌باشد: خلاصه یک یا چند متن، دربرداشتن اطلاعات مهم متون مبدأ و کوتاه بودن.

تحقیقات برای کشف دانش مهم و برجسته در داخل یک متن حول موضوع خلاصه‌سازی تک‌سنده می‌باشد[2]. این تحقیقات پیرامون دو دسته تولید خلاصه‌های چکیده‌ای[4] و استخراجی[5] بسط پیدا کرده‌اند. خلاصه استخراجی به معنی برگرداندن تعدادی از جملات متن به عنوان قسمت‌های مهم و خلاصه چکیده‌ای به معنای بیان دانش درونی یک متن به هر شکل ممکن می‌باشد[2].

در این مقاله ما به ارائه یک روش استخراجی خلاصه‌سازی برای یک متن، که تلفیقی از یک نظریه زبانشناسی به نام نظریه مرکزیت و برخی پارامترها و فرمول‌های آماری است می‌پردازیم. این تلفیق به منظورکاهش آثار مشکلات روش‌های کنونی خلاصه‌های استخراجی ارائه شده است. بیشتربودن طول جملات استخراجی از میانگین طول جملات متن، پراکندگی اطلاعات در سطح متن، تداخل اطلاعات جملات استخراج شده و عدم انسجام در خلاصه تولیدشده، از جمله این مشکلات می‌باشد. پارامترهای آماری به منظور حل مشکل اول و نظریه مرکزیت ایده اصلی برای حل باقی مشکلات می‌باشد. در بخش دوم این مقاله مروری بر ادبیات خلاصه‌سازی تک‌سنده استخراجی و نظریه مرکزیت خواهیم داشت.

در بخش سوم روش پیشنهادی و تعدادی از زیرالگوریتم‌های لازم برای پیش‌پردازش مورد بحث قرار می‌گیرند. در بخش چهارم نتایج مقایسه و ارزیابی روش پیشنهادی با تعدادی از سیستم‌های موجود ارائه شده‌است. در انتها نیز به جمع‌بندی مطالب ارائه‌شده می‌پردازیم.

## ۲- مرور ادبیات

### ۲-۱- خلاصه‌سازی تک‌سنده استخراجی

روش‌های زیادی برای خلاصه‌سازی یک متن ارائه شده‌است که به دلیل کثرت کارهای انجام شده به مهمترین آن‌ها اشاره می‌کنیم.

ادمونسون با ترکیب چهار پارامتر فرکانس کلمه، موقعیت جمله، توجه به کلمات نشانه‌ای مانند *این مهم است که، از آنجاکه و ....* و توجه به اسکلت جمله مانند کوتاه‌بودن و یا سربخش‌بودن آن، روش جدیدی برای خلاصه‌سازی ارائه کرد[2]. در یک پژوهش دیگر[3] از روش LSA به عنوان یک روش کلاسترینگ که از یک معیار وزن‌دهی لگاریتمی استفاده می‌کند، برای خلاصه‌سازی بهره برده شده است. در تحقیق دیگری با استفاده از یک شبکه عصبی برای پردازش داده‌های DUC2001-02 نتیجه‌گیری شده است که اولین جمله هر متن خبری، مهمترین جمله آن به شمار می‌رود[2]. در [4] روشی مبتنی بر استفاده از Lexical Chain[۶] ارائه شده‌است. همچنین [5] روشی مبتنی بر نظریه مرکزیت برای خلاصه‌سازی ارائه کرده‌است. در این روش با محاسبه CB هر جمله و اندازه‌گیری CBهای مشابه در متن، جملاتی که پرتکرارترین CBها را در خود داشته‌باشند، به عنوان خلاصه برگردانده می‌شوند.

## ۲-۲- نظریه مرکزیت

نظریه مرکزیت [6] یک مؤلفه از تئوری کلی تمرکز و انسجام گفتاری Grosz و Sidner است، که حول انسجام و برجستگی‌های محلی مطرح می‌شود. این نظریه در حال حاضر به وسیله [7] فرموله شده و توسط ملاک‌های تجربی مانند [8] تقویت شده‌است. از آنجایی‌که، این نظریه از توانایی بالقوه (تئوریک) خوبی برای تشخیص قسمت‌های مهم و منسجم یک متن برخوردار است، به‌عنوان ایده اصلی این پژوهش مورد استفاده قرار گرفته است. وابستگی میان جملات که انتقال (Transition) میان آن‌ها نامیده می‌شود، در نظریه مرکزیت به چهار کلاس اصلی Continue، Retain، Smooth-shift و Rough-shift دسته‌بندی می‌شود که در جدول (۱) قابل مشاهده هستند. برای تشخیص این انتقال‌های معنایی، پارامترهایی به این نظریه الصاق شده‌است، که یک ذات الگوریتمی به آن می‌بخشند. این پارامترها عبارتند از:

پارامتر $CF(U_n)$[۷]: لیستی از تمامی اسامی موجود در جمله n که بر اساس رابطه (۲) مرتب شده‌اند.

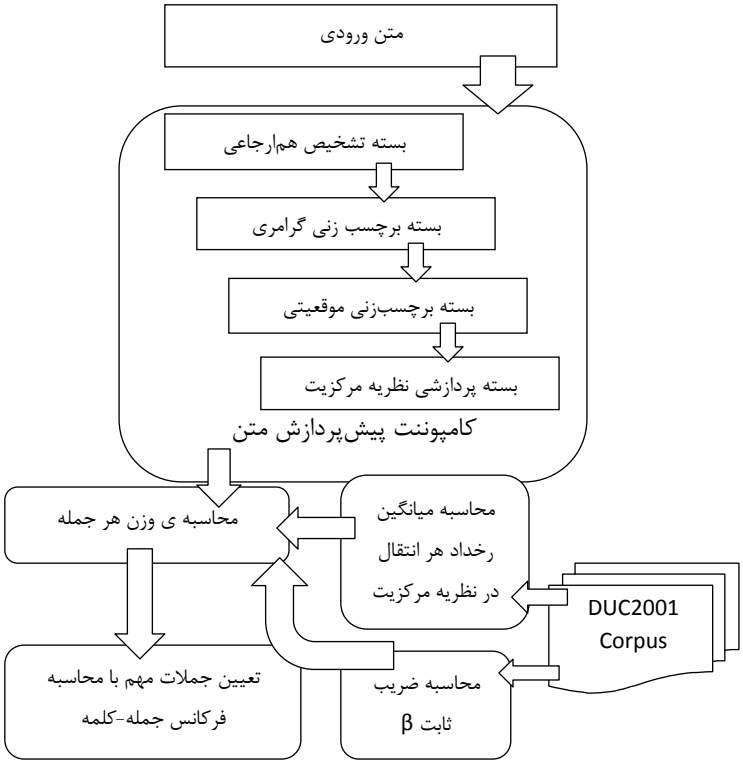
پارامتر $CP(U_n)$[۸]: پرارزشترین عنصر لیست CF می‌باشد که به اسم مرکز دارای ترجیح شناخته شده‌است.

پارامتر $CB(U_n)$[۹]: بالارتبه‌ترین اسم جمله است که برابر با پارامتر CP جمله n-1ام باشد.

در حال حاضر دو کاربرد مهم از نظریه مرکزیت در پردازش زبان طبیعی شامل کشف ارجاعات آنافوریک در یک متن [9] و مرتب‌سازی جملات [9] مورد استفاده و تحقیق می‌باشند.

| | Coherence: CB(Un) = CB($U_{n-1}$) Or CB($U_{n-1}$) ndef. | Coherence: CB(Un) ≠ CB($U_{n-1}$) |
|---|---|---|
| Salience: CB(Un) = CP(Un) | Continue | Smooth-Shift |
| Salience: CB(Un) ≠ CP(Un) | Retain | Rough-Shift |

**جدول شماره ۱- شرایط رخداد انتقال های ۴گانه ی نظریه ی مرکزیت بر اساس پارامترهای آن**




**شکل شماره (۱)- فلوچارت سیستم خلاصه‌ساز پیشنهادی**

## ۳- الگوریتم پیشنهادی

ایده اصلی این الگوریتم پس از مطالعه رفتار نظریه مرکزیت حول محور انسجام و وابستگی میان جملات یک متن بدست آمده‌است. این ایده شامل دو قسمت عمده‌است که منطبق بر دو اصل نظریه مرکزیت می‌باشند: با استفاده از اصل برجستگی عناصر مهم یک جمله تشخیص‌داده‌می‌شود، سپس با استفاده از اصل انسجام جملاتی که دارای عناصر مهم مشابه می‌باشند به عنوان جملات مهم و منسجم برای خلاصه هدف انتخاب می‌شوند. الگوریتم پیشنهادی شامل دو فاز مستقل می‌باشد. در فاز اول عملیات پیش‌پردازشی جهت آماده‌سازی متن برای کشف جملات مهم انجام می‌گیرد. در فاز دوم جملات بر اساس انتقال معنایی صورت‌گرفته میان آن‌ها

انتخاب‌شده و سپس خلاصه‌های تولیدشده ارزیابی می‌شوند. فلوچارت الگوریتم پیشنهادی در شکل شماره (۱) مشاهده می شود.

## ۳-۱- فاز پیش‌پردازش:

برای پردازش یک متن و کشف انتقال‌های میان جملات احتیاج به پیمودن گام‌هایی است که اصطلاحاً پیش‌پردازش متن نامیده می‌شوند. در ابتدا متن هدف باید توسط بسته پردازشی Corefrerence Resolution مورد پردازش قرار گیرد. در زبانشناسی یک هم‌ارجاعی زمانی رخ می‌دهد، که چندین عبارت در یک جمله یا متن به عبارت مشابهی اشاره می‌کنند. علت استفاده ازاین پیش پردازش این است که جایگزینی یک ضمیر یا یک عبارت اسمی با مرجع اسمی آن باعث به وجودآمدن درک و تفسیر درست و شفاف‌تری از متن و از بین‌رفتن ابهامات می‌شود.

برای انجام ماشینی این پردازش متنی ابزارهای آماده‌ای که به صورت آکادمیک تولید شده‌است، نیز وجود دارد. در این تحقیق از ابزار Cognitive Computation آزمایشگاه دانشگاه Illinois استفاده شده است که با دقت خوبی عملیات جانشینی مرجع را انجام می‌دهد[10]. هم‌اکنون تنها یک آمار مبنی بر میانگین دقت ۷۵٪ برای روش‌های جانشینی وجود دارد[11]. اما ابزار دانشگاه Illinois با استفاده از یک روش یادگیری ماشین دارای دقت ۸۰,۸٪ می‌باشد[12].

خروجی مرحله قبل پس از شکستن هر متن به جمله‌های تشکیل-دهنده‌اش، بر اساس محل قرارگرفتن نقطه به عنوان ورودی مرحله برچسب زنی نقش معنایی (Semantic Role Labeling) استفاده شده‌است. این ابزار به دنبال تعیین نقش معنایی هر عنصر جمله در ارتباط با فعل آن جمله می‌باشد. در حین انجام این عملیات نقش گرامری آن عنصر از قبیل فاعلی، مفعولی (مستقیم و غیر-مستقیم)، انواع قیود و ... نیز مشخص می‌گردد.

در این مرحله ما از ابزار گروه Cognitive Computing دانشگاه Illinois استفاده شده‌است که از پارسر Charniak برای برچسب-زنی به عناصر جمله استفاده می‌کند[10]. سپس برای هر متن یک ماتریس تشکیل داده‌می‌شود. مطابق فرمول (۱) سطرهای این ماتریس شامل جملات متن و ستون‌های ماتریس شامل نقش‌های گرامری مورد توجه در تحقیقات می‌باشد.

حال بر روی هر خانه از ماتریس بالا عملیات برچسب‌زنی موقعیتی (POS Tagging) انجام می‌شود. البته ورودی ابزار POS Tagger کل متن‌می باشد. اما پس از برچسب‌زنی کلمه‌ای که به عنوان اسم اصلی در هر عبارت اسمی در هر خانه‌ی ماتریس برچسب می‌خورد تعیین‌شده و مابقی عبارت اسمی حذف می‌شوند.

(1) Sentence-Term Matrix:

$$\begin{array}{c} \\ \text{Sentences} \end{array} \begin{bmatrix} Term11 & .. & .. & .. & Term1N \\ .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. \\ TermM1 & .. & .. & .. & TermMN \end{bmatrix}_{M \times N} \quad \text{Terms}$$

### ۳-۱-۱- تشکیل پارامترهای نظریه مرکزیت و تعیین انتقال‌های میان جملات

در این فاز پارامترهای نظریه‌ی مرکزیت برای هر جمله از هر متن تعیین می‌گردد. بر اساس تعریف نظریه‌ی مرکزیت ترتیب اسامی در یک جمله مطابق رابطه (۲) می‌باشد.

(۲) فاعل > مفعول غیر مستقیم > مفعول مستقیم > قید

همانطور که ذکر گردید در ماتریس جمله–لغت که در مراحل گذشته تشکیل شد، ستون‌ها که نشان‌دهنده لغات می‌باشند، توسط نقش گرامریشان برچسب‌خورده‌اند. در نتیجه تا همین مرحله پارامترهای CF و CP برای هر جمله تشکیل شده است. شبه کد مورد نظر برای تعیین پارامتر CB مطابق شکل(۲) می‌باشد.

```
For sentence number (n)
        Tag the most important term as CP
        X = CP(n);
End;
CB(n+1) = NULL;
For each term in sentence number (n+1)
        If(Sentence-Term((n+1),i) = X)
                CB(n+1) = X;
        End;
End;
```

**شکل (۲) – شبه کد محاسبه ی پارامتر CB از نظریه ی مرکزیت برای هر جمله از ماتریس Sentence - Term**

حال قوانین مربوط به ۴ انتقال معنایی مورد توافق در نظریه مرکزیت بر روی ماتریس کامل برچسب‌خورده جمله–لغت اعمال می‌شود. شبه کد این کار در شکل (۳) ارائه شده‌است. ماتریس جدید، یک ماتریس با ابعاد m*(n+1) می باشد که ستون n+1 ام نشان‌دهنده انتقال معنایی میان جمله جاری و جمله قبلی می‌باشد که در رابطه شماره (۷) دیده می‌شود.

```
For each sentence
   If(CB(m) = CB(m-1) Or CB(m-1) undef)
      If(CB(m) = CP(m))
        Sentence-Term(m,n+1) = Continue;
      Else Sentence-Term(m,n+1) = Retain;
   If(CB(m) ≠ CB(m-1))
      If(CB(m) = CP(m))
         Sentence-Term(m,n+1) = Smooth-shift;
      Else Sentence-Term(m,n+1) = Rough-shift;
End;
```

**شکل (۳) – شبه کد محاسبه ی انتقال‌های معنایی ۴گانه میان هر دو سطر متوالی از ماتریس Sentence - Term**

## ۳-۲- فاز انتخاب جملات

در این تحقیق برای انتخاب جملات مهم گام‌هایی پیموده شده‌است که به ترتیب به آن‌ها اشاره خواهد شد.

### ۳-۲-۱- وزن‌دهی به جملات بر اساس انتقال انجام‌شده میان دو جمله متوالی

در این تحقیق بر اساس یک حدس اولیه که مبتنی بر ساختار انتقال‌های معنایی می‌باشد، برخی از جملات بر حسب انتقال معنایی رخ داده میان آن‌ها به عنوان جملات مهم انتخاب شده‌اند. انتقال‌های معنایی در نظریه مرکزیت، طبق آنچه نشان داده‌شد، در حقیقت ارتباط میان کلمات دو جمله متوالی می‌باشند. بر این اساس هر بار یکی از انتقال‌های معنایی مورد توجه قرار گرفته‌است و به‌ترتیب هر کدام از این انتقال‌ها مورد ارزیابی قرارگرفته‌اند. برای هر کدام از انتقال‌ها میانگین رخداد آن‌ها میان جملات یک متن در کل پیکره مورد ارزیابی به دست‌آمده‌است که مطابق جدول شماره (۲) قابل مشاهده است. بر این اساس نوع امتیازدهی به این انتقال‌ها برای انتخاب جملات مهم مطابق رابطه (3) می‌باشد.

$$TF = \left(\frac{f_{transition,i}}{\sum_k f_{transition,k}}\right) Log_2\left(\frac{|D|}{|\{d:transition \in d\}|}\right) \quad (3)$$

مقدار به دست‌آمده برای هر انتقال، یک مقدار عددی ثابت است که از محاسبه مقادیر گفته‌شده بر روی DUC2001 به عنوان مجموعه داده یادگیری به دست آمده‌است. بررسی‌ها و نتایج به‌دست‌آمده، دقت مناسب این ضریب محاسبه‌شده را نشان می‌دهند.

| Transition | Occur Average in DUC2001 |
|---|---|
| Continue | 0.15 |
| Retain | 0.24 |
| Smooth-shift | 0.19 |
| Rough-shift | 0.42 |

**جدول شماره (۲)- مقدار میانگین نسبی رخداد هر انتقال معنایی در هر متن**

#### ۳-۲-۱-۱ وزن‌دهی به انتقال‌ها بر اساس فرکانس تکرار در متن

بر اساس یک شهود می‌توان حدس زد که میزان تکرار بالای یک انتقال میان جملات یک متن از اعتبار آن انتقال می‌کاهد. به همین دلیل با محاسبه فرکانس نسبی تکرار هر انتقال در هر متن و نرمال-کردن آن این مشکل را برطرف خواهدشد. رابطه (4) فرمول محاسبه این فرکانس نسبی نرمال‌شده را نشان می‌دهد.

$$TW_n = \beta * (MT_n - MA) / VA * TF \quad (4)$$

Where $VA := {}^1/_4 \sum_{i=1}^{4} (MT_i - MA)^2$

And $MT_n := \frac{\sum_{i=1}^{number\ of\ this\ transition\ n} 1}{\sum_{k=1}^{number\ of\ transitions} 1}$

And $MA := {}^1/_4 \sum_{j=1}^{4} \frac{\sum_{i=1}^{number\ of\ this\ transition\ j} 1}{\sum_{k=1}^{number\ of\ transitions} 1}$

بر اساس آزمایش‌های صورت‌گرفته مقدار ضریب $\beta$ برای هر انتقال مطابق جدول شماره (۳) به‌دست آمده‌است.

| Transition | β Factor |
|---|---|
| Continue | 2.6 |
| Retain | 1 |
| Smooth-shift | 1.8 |
| Rough-shift | 0.85 |

**جدول شماره ی (۳)- مقادیر مختلف ضریب β در رابطه ی شماره ی (4)**

### ۳-۲-۲- محاسبه فرکانس نسبی کلمه-جمله و نرمال کردن آن

حجم جملات استخراج‌شده بر حسب انتقال‌ها در برخی از موارد بیشتر از حجم لازم برای خلاصه می‌باشد. برای کاهش حجم گام‌های زیر را مورد استفاده قرار داده خواهد شد.

#### ۳-۲-۲-۱- وزن‌دهی به کلمات

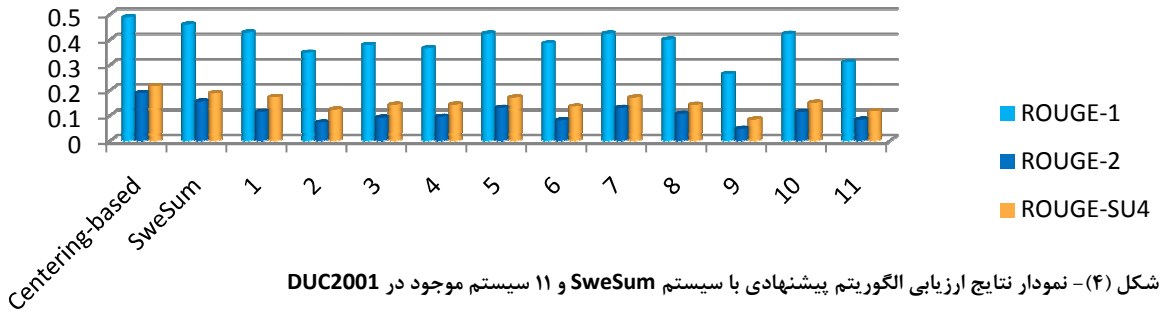
برای هر کلمه در جمله بر اساس رابطه (5) مقدار TF-ISF محاسبه می‌گردد.

$$(TF * ISF)_{j,i} = \left(\frac{tf_{j,i}}{\sum_k tf_{j,k}}\right) Log_2\left(\frac{|D|}{|\{d:t_i \in d\}|}\right) \quad (5)$$

در این رابطه $tf_{j,i}$ مقدار تکرار یک کلمه در یک جمله، $|D|$ تعداد کل جملات یک متن و $|\{d: t_i \in d\}|$ تعداد کل جملاتی است که کلمه مورد نظردر آن‌ها دیده شده‌است. حال با استفاده از رابطه (6) میانگین وزن کلمات یک جمله را محاسبه و آنرا نرمال می‌کنیم.

$$SMT_j := \frac{\sum_{i=1}^{n} NTFPS_{j,i}}{\sum_{i=1}^{n} NTFPS_{j,i}>0} \quad (6)$$

Where $\text{NTFPS}_{j,i} := \left(\left(\frac{tf_{j,i}}{\sum_k tf_{j,k}}\right) - MTPS\right) / VPS$




**شکل (۴)- نمودار نتایج ارزیابی الگوریتم پیشنهادی با سیستم SweSum و ۱۱ سیستم موجود در DUC2001**

And $$MTPS_j := \frac{\sum_{i=1}^{n}\left(\frac{tf_{j,i}}{\sum_k tf_{j,k}}\right)}{\sum_{i=1}^{n} tf_{j,i}>0}$$

And $$VPS_j := \frac{\sum_{i=1}^{n}\left(\left(\frac{tf_{j,i}}{\sum_k tf_{j,k}}\right)-\sum_{i=1}^{n}\left(\frac{tf_{j,i}}{\sum_k tf_{j,k}}\right)\right)^2}{\sum_{i=1}^{n} tf_{j,i}>0}$$

حال مقدار وزن به‌دست‌آمده برای هر جمله را در وزن به‌دست‌آمده از رابطه (4) ضرب کرده و در خانه ی n+2 ماتریس رابطه (7) قرار می‌دهیم. از این پس ماتریس جمله–کلمه یک ماتریس m × (n+2) خواهد بود. سپس ماتریس جمله–کلمه را بر حسب ستون n+2 آن مرتب کرده و جملات را بر اساس میزان فشردگی مورد نظر انتخاب می‌کنیم.

Terms

$$\text{Sentences}\begin{bmatrix} Term11 & \dots & \dots & \dots & Term1N & Transition & Weight \\ \vdots & \ddots & \dots & \dots & \vdots & \vdots & \vdots \\ \vdots & \dots & \ddots & \dots & \vdots & \vdots & \vdots \\ \vdots & \dots & \dots & \ddots & \vdots & \vdots & \vdots \\ TermM1 & \dots & \dots & \dots & TermMN & Transition & Weight \end{bmatrix}_{M \times (N+2)}$$

## ۴- ارزیابی کارایی

### ۴-۱- مجموعه داده‌ها

همانطور که پیشتر گفته شد در ارزیابی و پیاده سازی این الگوریتم از مجموعه داده‌هایی با عنوان DUC2001 استفاده شده است. ساختار کلی این مجموعه ها به صورت تعدادی کلاستر است، که در هر کلاستر چندین متن خبری با یک موضوع از خبرگزاری های معتبر موجود می باشد. به این جهت برای هر متن و هر کلاستر خلاصه های انسانی چکیده ای در چند حجم مختلف موجود می باشد. در این مقاله تمرکز بر روی خلاصه های تک سنده با حجم ۱۰۰ کلمه‌ای می باشد. DUC2001 شامل 600 سند می باشد که در ۶۰ کلاستر دسته بندی شده اند. در این تحقیق برای هر متن الگوریتم پیشنهادی پیاده سازی شده و سپس خلاصه ی استخراج شده با خلاصه های سیستم های خلاصه ساز ماشینی معرفی شده در DUC2001 مقایسه شده است. همچنین برای هر متن یک خلاصه توسط ابزار SweSum که یک ابزار خلاصه ساز آنلاین تجاری و مؤفق است تهیه کرده ایم. در این ارزیابی نتیجه با خلاصه ی ایجاد شده توسط این ابزار نیز مقایسه شده است. برای ارزیابی از ابزار ROUGE[۱۰] استفاده نمودیم. در سال های اخیر، در مقالات مختلف از این ابزار برای ارزیابی به دفعات استفاده شده است. سیستم های موجود در DUC هم با این ابزار ارزیابی شده اند.

### ۴-۲- روش ارزیابی

به منظور ارزیابی ابتدا معیار F-measure که ترکیبی از دو معیار recall و Precision می باشد، مورد توجه قرار گرفته‌است. میانگین معیار F-measure برای ۱۲ سیستم ماشینی به طور جداگانه و سیستم پیشنهادی، به ازای تمامی 600 متن موجود در DUC2001 در سه معیار ROUGE1، ROUGE2 و ROUGESU4 در شکل شماره (4) دیده می شود. همانطور که مشاهده می شود در هر سه معیار فوق الذکر عملکرد روش پیشنهادی بهبود یافته است.

## نتیجه گیری

در این مقاله با در نظر گرفتن مشکلات ۴ گانه ی روش های موجود برای خلاصه سازی تک سنده ی استخراجی، روش جدیدی با تلفیق یک تئوری زبانشناسی به نام نظریه مرکزیت و اتکا بر پارامترهای آماری متن، توانستیم به نتایج بهتری نسبت به روش های کنونی دست یابیم. به نظر می رسد که توانایی های تئوری مرکزیت در تشخیص انسجام در متن بالاتر از نتایج به دست آمده باشد.

یکی از حوزه های جذاب که می توان از این تئوری در آن بهره برد، خلاصه سازی چند سندی، استخراج مفاهیم از داخل مجموعه ای از متون و .... می باشد. دست یافتن به الگوهایی از انتقال های معنا دار در مجموعه ای از متون از اهداف آینده می باشد.

## مراجع

[1] - Radev, D. R., Hovy, E., and McKeown, K. (2002). Introduction to the special issue on summarization. Computational Linguistics., 28(4):399-408.
[2] - D. Das and A. Martins. A Survey on Automatic Text Summarization. Literature Survey for the Language and Statistics II Course at CMU, 2007.
[3] - Gong, Y., & Liu, X. Generic text summarization using relevance measure and latent semantic analysis.Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval,SIGIR`01, New Orleans, Louisiana,United States, 2001.
[4] - Barzilay, R. and Elhadad, M. (1997). Using lexical chains for text summarization. In Proceedings ISTS'97.
[5] - Hoffman, B. (1996). Summarization: an Application for NL Generation, Proceeding of International workshop on NL Generation,1996.
[6] - Joshi, A. K. and Weinstein, S. (1981). Control of inference: Role of some aspects of discourse structure–centering. In Proc. International Joint Conference on Artificial Intelligence, pages 435– 439.
[7]-Chafe,W.(1976).Givenness,contrastiveness, efiniteness, subjects, and topics. In Li, C., editor, Subject and Topic, pages 25–76. Academic Press, New York.
[8] - Kintsch, W. and van Dijk, T. (1978). Towards a model of discourse comprehension and production. Psychological Review, 85:363–394.
[9] - Milan, T. (2009). Extending Centering Theory for the Measure of Entity Coherence.MSc thesis,Simon Fraser.
[10]- http://cogcomp.cs.illinois.edu/
[11] - **(Elango, 2005)** Pradheep Elango. Coreference Resolution: A Survey. Technical Report, University of Wisconsin Madison, 2005. PDF
[12] - Eric Bengtson , Dan Roth, Understanding the value of features for coreference resolution,Proceedings of the Conference on Empirical Methods in Natural Language Processing, October 25-27, 2008, Honolulu, Hawaii

---

[1] Semantic Nets
[2] Centering Theory
[3] Part of Speech Tagger
[4] Abstractive
[5] Extractive
[6] زنجیره ی لغوی
[7] Forward looking Center
[8] Prefer Center
[9] Backward looking Center
[10] - Recall-Oriented Understudy for Gisting Evaluation