

امروزه با توجه به حجم گسترده اطلاعات موجود، کمبود زمان بررسی و پردازش مطالب و همچنین عدم وجود یک سیستم خلاصه ساز کارآمد برای خلاصه سازی حجم انبوه کتب، مقالات و اخبار در سطح وب نیاز به یک سامانه قدرتمند جهت خلاصه سازی متون احساس می‌شود. با توجه به اینکه بر روی موضوع خلاصه سازی تحقیقات زیادی انجام شده و مقالات زیادی در مورد آن منتشر گردیده، اما همچنان ضعف‌های زیادی در این زمینه وجود داشته و تا رسیدن به سیستم کارآمدی که بتواند این عمل را مشابه یک عامل هوشمند انسانی انجام دهد فاصله زیادی باقی مانده است. مشکل خلاصه سازی در زبان فارسی به مراتب بیشتر از زبان‌های دیگر است. پیچیدگی‌های زبانی موجود و عدم وجود ابزارهای با دقت مناسب برای کار با زبان فارسی از جمله مشکلات فعلی پیش رو در پردازش متون فارسی است. از این رو بررسی عملیات و روش‌های انجام شده بر روی سایر زبان‌ها، توجه به حوزه معنایی و استفاده از این رابطه‌های معنایی، بکارگیری ابزارهایی مانند گراف‌ها و روش‌های آماری می‌توانند کمک شایانی به این مهم بکنند.

توجه و علاقه به خلاصه سازی خوکار متن، اولین بار حدود دهه پنجاه به وجود آمد. اولین فعالیت‌ها در این زمان توسط فردی به نام Luhn شروع شد. در ابتدا اساس کار او، یافتن کلمات با بیشترین تکرار بود. از نظر وی کلمات با فرکانس تکرار بیشتر در یک متن مهمتر از سایر کلمات بوده و جملاتی که تعداد بیشتری از این کلمات را دارند مهمترین بخش‌های متن بوده و باید در متن خلاصه شده قرار بگیرند. البته روش اولیه ارائه شده توسط وی خطای زیادی داشت که بعدها توسط خود او اصلاحاتی بر روی آن انجام شد. به عنوان مثال برخی افعال و حروف اضافه دارای فرکانس بالایی در تمامی متون بوده اما حاوی اطلاعات مهمی نبوده و در ایده قبل او دارای اهمیت محسوب می‌شدند. در اصلاحات بعدی حروف اضافه و برخی از کلمات پرتکرار حذف شده و در الگوریتم استخراجی شرکت داده نشدند. گرچه که روش پیشنهادی وی دارای دقت خوبی نبود اما به دلیل اینکه پایه گذار اصلی خلاصه سازی متن بود بسیار مورد ستایش قرار گرفت. از آن به بعد خلاصه سازی متن یکی از حوزه‌های مهم تحقیقاتی در پردازش زبان طبیعی در نظر گرفته شده و تحقیقات زیادی بر روی آن انجام شده است. از آن تاریخ به بعد روش‌های زیادی به همراه ابزارهایی قدرتمند به کار گرفته شده تا بتوانند پردازش متن را مانند آنچه که در مغز انسان انجام می‌شود را شبیه سازی کنند. ادmondson (Edmondson) از جمله کسانی بود که بعد از وی از سایر ویژگی‌های موجود در متن برای ایجاد خلاصه‌های بهینه استفاده کرد. وی برای مشخص کردن ارزش هر جمله از ترکیب خطی این ویژگی‌ها استفاده کرد. روش‌های ارائه شده برای تعیین وزن جملات به صورت زیر بود:

- **روش نشانه:** این روش مبتنی بر ارتباط جملات و کلمات بود، به این صورت که وجود یا عدم وجود برخی از کلمات در جمله با استفاده از یک فرهنگ لغات محاسبه می‌شد.
- **روش عنوان:** در این روش وزن جمله براساس تمام کلماتی که در عنوان و زیر عنوان‌های متن وجود داشت محاسبه می‌شد.
- **روش مکان:** در این روش مکان هر جمله با یک بخش از متن بررسی می‌شد و احتمال ارتباط بین آنها مورد نظر قرار می‌گرفت.

نتایج نشان داد که استفاده از سه روش فوق باعث شد که خلاصه سازی خودکار متن شباهت زیادی به خلاصه سازی انسان داشته باشد. از سال 1995 به بعد معیارهای دیگری نیز برای تعیین قسمت‌های مهم یک متن معرفی شدند.