

کشف تقلب با استفاده از روش مبتنی بر درخت Trie

نویسندگان :

دکتر طالب پور ، محمد شیرزادی، زهرا امین الرعایا

زمستان ۱۳۹۵

فهرست مطالب

✓ سرقت ادبی چیست؟

✓ استلزام اتوماسیون کشف تقلب

✓ تشریح روش مبتنی بر درخت Trie

سرقت ادبی چیست؟

✓ سرقت علمی- ادبی استفاده از ایده‌ها و عبارات دیگران، به عنوان ایده و عبارات خویش و در واقع نوعی دستبرد فکری یا ایده دزدی است.

✓ سرقت علمی- ادبی می‌تواند شامل طیف وسیعی از دستبردهای آگاهانه تا کپی کردن اتفاقی مطالب دیگران باشد.

• استلزام اتوماسیون کشف تقلب

✓ رشد روز افزون حجم محتوای متنی.

✓ امکان بررسی این حجم از داده توسط یک شخص (تولید کننده محتوا یا تایید کننده محتوا (کار بسیار دشواری است.

✓ نیاز به وجود یک سیستم اتوماتیک جهت آنالیز خودکار محتوا می باشد.

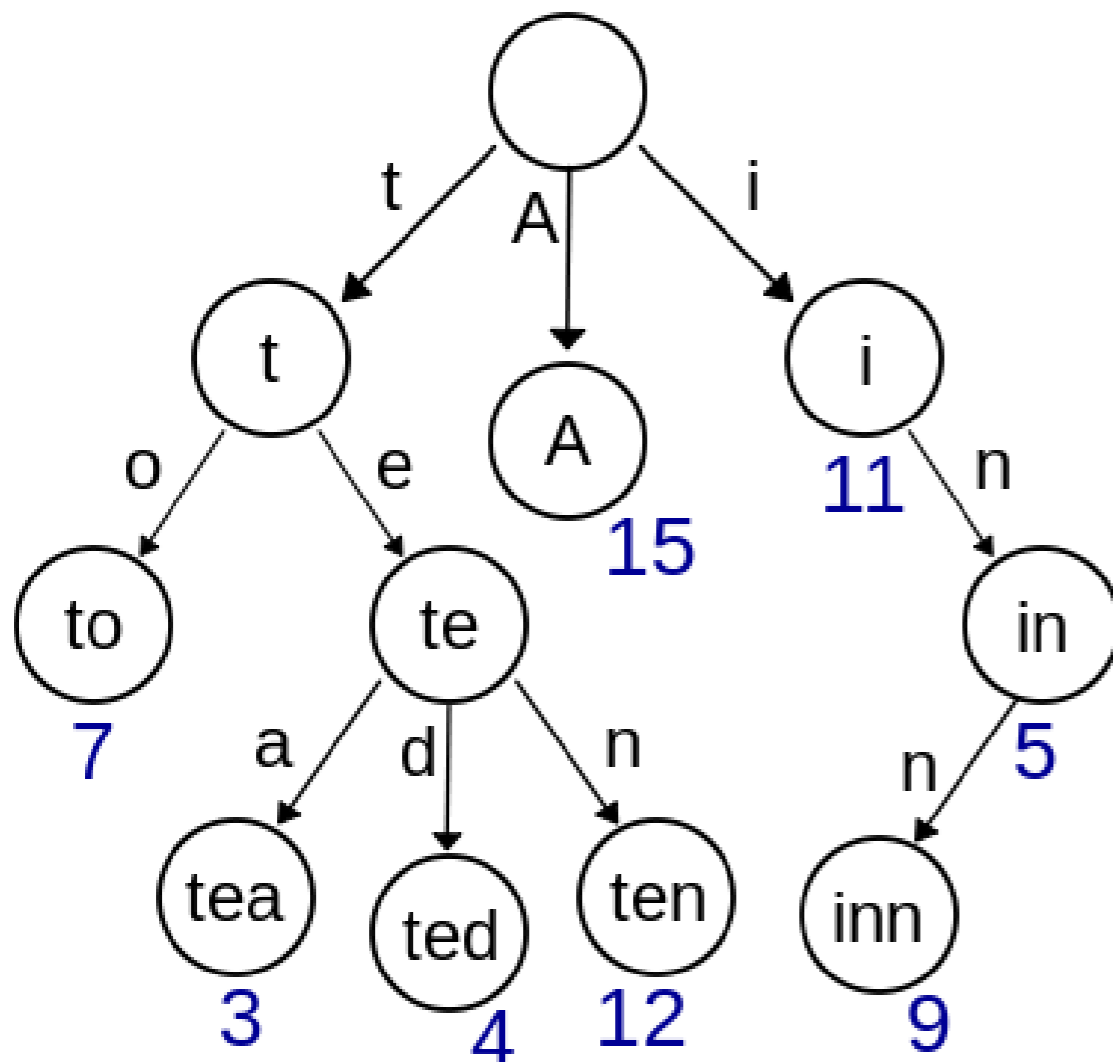
• تشریح روش مبتنی بر درخت Trie

✓ آماده سازی اطلاعات

- Text Cleansing & Normalization
- Tokenization & POS Tagging
- Removing Stop words and frequent words
- Acquiring words synsets

• تشریح روش مبتنی بر درخت Trie

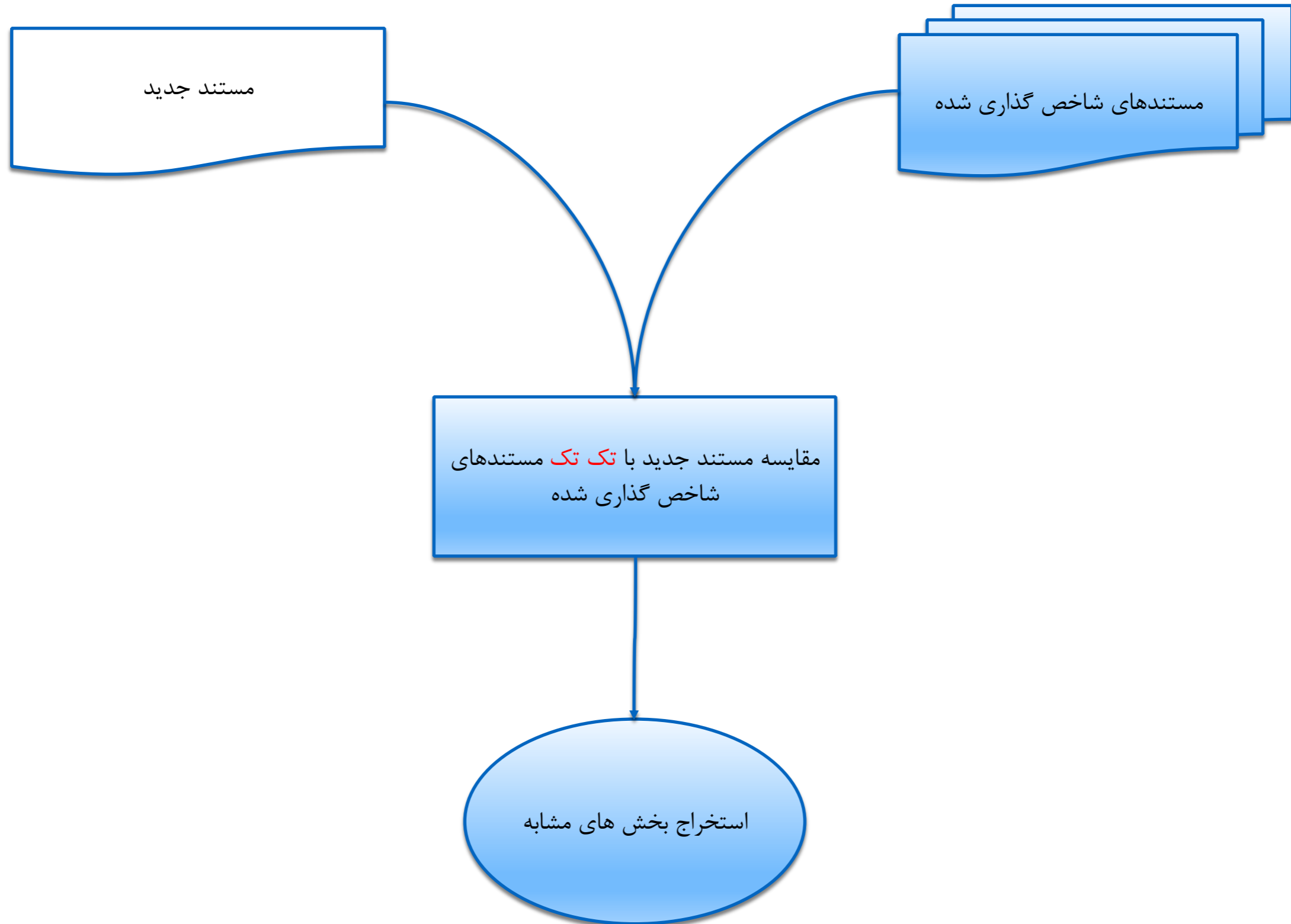
- ایجاد درخت Trie از روی مستندهای ذخیره شده
- درخت شامل لیست تمام کلمات با موقعیت مکانی آنها در مستند می باشد.



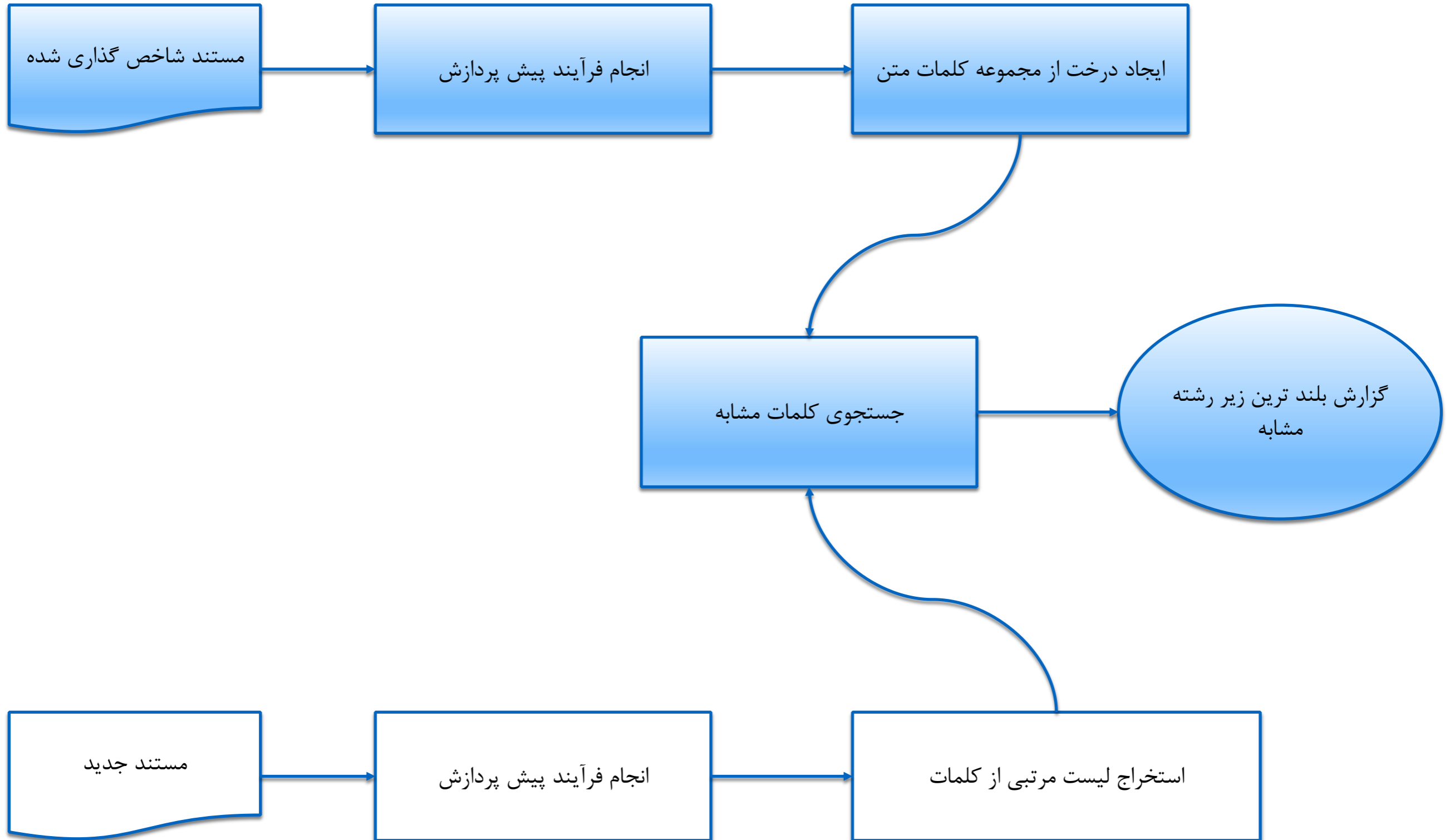
ten

- ✓ Occurrence Positions
- ✓ Front words positions
- ✓ Rear word positions

• تشریح روش مبتنی بر درخت Trie



• تشریح روش مبتنی بر درخت Trie



پایان