

Text mining an Effective Tool for Documenting and Completing Oral History interviews

Peymaneh Salehi

Author and Researcher of Oral History

Abstract: Oral history, as one of the new methods of collecting information, has gone through a path full of ups and downs to establish itself as a research tool for historical studies. This research method has a narrative feature and recounts events based on the insights and hearings of observers of historical periods. In the meantime, documenting, verifying and interpreting the texts obtained from the interviews is very important. The purpose of this paper is to emphasize the need to use practical information processing tools such as text mining and discover semantic relationships in order to complete the sources of oral history. The need to use modern information processing tools in the process of oral history has been explained and analyzed with reference to library resources and methods of using it. The use of these technologies makes it easier to supplement the data and improve the content of the interview and ultimately turn it into a research-based and citationable work and brings the interview closer to its goals.

Key Words: Oral history, Interview, Documentation, Text mining, Semantic relations.

متن کاوی ابزاری کارآمد به منظور مستندسازی و تکمیل مصاحبه‌های تاریخ شفاهی

پیمانہ صالحی*

نویسنده و پژوهشگر تاریخ شفاهی، کارشناس ارشد زبان و ادبیات فارسی

تاریخ دریافت: ۱۴۰۰/۰۵/۲۹ تاریخ پذیرش: ۱۴۰۰/۰۷/۱۷

چکیده

هدف: اهمیت به کارگیری ابزار متن کاوی به منظور مستندسازی منابع تاریخ شفاهی و کشف روابط معنایی در مصاحبه‌ها برای تکمیل داده‌های گردآوری شده و تبدیل مصاحبه‌ها به آثاری پژوهشی، با قابلیت استناد.

روش: لزوم بهره‌مندی از ابزارهای نوین پردازش اطلاعات در فرآیند تاریخ شفاهی با استناد به منابع کتابخانه‌ای تبیین شده است و روش‌های استفاده از آن بازنمایی و تجزیه و تحلیل شده‌اند.

یافته‌ها: بهره‌گیری از فناوری‌های نوین همچون متن‌کاوی و کشف روابط معنایی، در راستای رفع ابهام و راستی آزمایی مصاحبه‌های تاریخ شفاهی به دلیل دقت، جامعیت و سهولت در استفاده، موجب ارتقای داده‌های حاصل از مصاحبه‌ها می‌شود. بهره‌مندی از این ابزارها در زمان انجام مصاحبه و پس از آن در هنگام تدوین و ویراستاری متن، کاربرد دارد. هرچه پژوهشگران تاریخ شفاهی به مستندسازی، گویاسازی و راستی‌آزمایی مصاحبه‌ها در مراحل مختلف پژوهش و با تمرکز بر قالب‌های نوین اطلاعات، بیشتر توجه کنند، متن نهایی به اهداف انجام مصاحبه، یعنی تولید داده‌های جدید در حوزه موضوعی خاص، نزدیکتر می‌شود.

واژه‌های کلیدی: تاریخ شفاهی، مصاحبه، مستندسازی، متن کاوی، روابط معنایی.

مقدمه

گسترش پایگاه داده‌ها در عصر ارتباطات، موجب افزایش تقاضا به منظور تبدیل داده‌ها به دانش مقید شده است. برای دسترسی به این دانش باید از ابزارهای قدرتمند بهره جست. متن‌کاوی^۱ یا استخراج دانش از پایگاه‌های متنی، یکی از همین ابزارها است. کشف و شناسایی دانش از متن^۲ به این معنا است که مفاهیم صریح و ضمنی و روابط معنایی بین داده‌ها با استفاده از تکنیک‌های پردازش زبان‌های طبیعی^۳ استخراج شود.

انجام مصاحبه تاریخ شفاهی یکی از روش‌های ثبت رویدادهای گذشته است؛ البته مزیت پژوهش بر پایه مصاحبه در این است که اطلاعات بی‌واسطه به دست می‌آیند و محقق را قادر می‌سازد تا تفاوت‌های ظریفی را دریابد که درباره مطالعه گذشته در خلال بازتاب تاریخ بروز می‌کند. اگر پژوهش تاریخ شفاهی به درستی انجام پذیرد، به پژوهشگران امکان می‌دهد تا به لایه‌های پنهان اسناد مکتوب پی ببرند و علاوه بر درک آنچه در گذشته روی داده، بدانند آن رویداد چگونه رخ داده است.

مستندسازی، گویاسازی و راستی‌آزمایی مصاحبه‌های تاریخ شفاهی در زمان انجام مصاحبه و پس از آن طی پروسه تدوین متن، یکی از بخش‌های حائز اهمیت در این روش پژوهشی قلمداد می‌شود. مصاحبه پس از عبور از این صافی‌های کیفی، به اثری با قابلیت استناد تبدیل می‌شود. در این میان، بهره‌مندی از ابزارهای نوین با هدف تکمیل داده‌های حاصل از تاریخ شفاهی، ضمن ایجاد سرعت و سهولت در این فرایند، با امکاناتی که در اختیار پژوهشگران قرار می‌دهد، موجب ارتقای سطح اطلاعات می‌شود و یاری‌رسان تاریخ شفاهی‌نگاران خواهد بود.

پژوهش حاضر در پی پاسخ به پرسش‌های زیر است:

۱. چه روش‌هایی برای کشف دانش و پردازش زبان‌های طبیعی رواج بیشتری دارد؟
۲. فرآیند متن‌کاوی با بهره‌مندی از کدام تکنیک‌ها به سطح بالاتری از کاربرد می‌رسد؟

۳. استخراج روابط معنایی در منابع، چگونه موجب سهولت و دقت در مستندسازی و گویاسازی مصاحبه‌ها را فراهم می‌آورد؟

۴. کاربرد متن‌کاوی به منظور تکمیل مصاحبه‌های تاریخ شفاهی، چه مرحله‌ای را شامل می‌شود؟
در سال‌های اخیر، آثاری در حوزه متن‌کاوی، داده‌کاوی و وب معنایی و کاربرد این فناوری‌ها در علوم مختلف تألیف و ترجمه شده‌اند که تعدادی از آنها ذکر می‌شود:

امیر علی‌خانزاده (۱۳۸۵) با ترجمه کتاب *داده‌کاوی*، اثر مهمد کانتاردزیک^۴ ضمن بررسی مفاهیم آن، کاربردهای داده‌کاوی را تجزیه و تحلیل کرده است.

کیوان معقولی و مرتضی زنگنه سروش (۱۳۹۱) با ترجمه کتاب *اکتشاف دانش اثر دانیل تی لاروز* ضمن تشریح مفاهیم پایه‌ای داده‌کاوی، آن را بستر پیشرفت علوم مختلف معرفی کرده‌اند.

نوید شیدایی (۱۳۹۱) با نگارش پایان‌نامه‌ای با عنوان *متن‌کاوی متون فارسی*، نحوه پیش‌پردازش و دسته‌بندی مقالات خبری را تشریح کرده است.

سودابه پارسا (۱۳۹۳) با تألیف پایان‌نامه‌ای با عنوان *توسعه و بهبود روش‌های متن‌کاوی در کاربردهای فارسی*، دشواری‌های این فناوری را در متون فارسی، تحلیل و با ذکر مطالعات موردی راهکارهایی برای حل آن مطرح کرده است.

بابک تیمورپور و حیدر نجفی (۱۳۹۴) با تألیف کتاب *داده‌کاوی با R به همراه متن‌کاوی و تحلیل شبکه‌های اجتماعی*، کاربرد متن‌کاوی در شبکه‌های اجتماعی را تجزیه و تحلیل کرده‌اند.

مریم علی‌کوثری (۱۳۹۶) با ترجمه کتاب *داده‌متن‌کاوی اثر چارو سی آگروال و چنگ‌شیانگ جای*^۵ به کاربرد داده‌کاوی در علوم مختلف به تفکیک توجه داشته است.

عاطفه قهرمانی‌فر (۱۳۹۶) با نگارش پایان‌نامه‌ای با عنوان *استخراج کلمات کلیدی با استفاده از تکنیک متن‌کاوی*، به پردازش زبان‌های طبیعی و شباهت‌سنجی اسناد برای پیدایی کلیدواژه‌ها اشاره کرده است.

^۴ Mehmed Kantardzic.

^۵ Charo C Aggarwal & Chengxiang Zhai.

^۱ Text Mining.

^۲ Knowledge- Discovery in Text (KDT).

^۳ Natural Language Processing.

می‌شود. پس از آن، مفاهیم داده‌کاوی، متن‌کاوی، وب‌کاوی^۶ و کلمات کلیدی، تشریح و تکنیک‌های آنها تبیین می‌شوند تا به‌کارگیری این ابزارها در زمان انجام مصاحبه و نیز در هنگام تدوین و ویرایش متن بررسی شود.

فرض کنید دو واژه «ابركوه» و «رُستنی» را به شما بدهند. استفاده از متن‌کاوی، ترکیبات زیر را از این دو واژه می‌سازد:

- در مرحله اول «شهر، یزد، گیاه و درخت»؛

- در مرحله دوم «شهر کهن و درخت سرو»؛

- در مرحله سوم «سرو چهار هزار ساله ابرکوه»؛

- در مراحل بعدی، مواردی نظیر حجم آب برای آبیاری درخت سرو، انواع رُستنی‌های مناطق گرم و خشک، مشخصات جغرافیایی ابرکوه، بررسی‌های علمی انجام‌شده در این منطقه، وضعیت یزد و ابرکوه در دوره‌های تاریخی و آداب و رسوم و فرهنگ این مناطق بررسی می‌شوند. به عبارت روشن‌تر، با کشف اطلاعات نهفته در منابع و دسته‌بندی روابط پیچیده میان داده‌ها، نتایج مهمی به دست می‌آیند. همچنین، سازمان‌های مرتبط با این موارد، از نقصان اطلاعاتی در حوزه‌های مرتبط با آن آگاه می‌شوند و به چنین پرسش‌هایی پاسخ می‌دهند:

• چه افرادی این موضوع را برای نخستین بار در رسانه‌ها مطرح کردند؟

• براساس موارد مشابه، لازم است چه اقداماتی در آینده باید برای این درخت انجام شود؟

• امکانات طبیعی ابرکوه چه مواردی را شامل می‌شود؟ بنابراین، موارد فوق، بخشی از قابلیت‌های متن‌کاوی را در استخراج دانش و کشف روابط معنایی آشکار می‌سازد.

تافری^۷ داده‌کاوی را این‌گونه تعریف می‌کند: «داده‌کاوی یا کشف دانش، استخراج اطلاعات از داده‌ها و استفاده از الگوریتم‌های خاص برای این فرایند است. هدف از کشف دانش در پایگاه داده‌ها، بازیابی الگوهای پنهان و نمایان‌شدن پیوستگی و ارتباط بین داده‌ها است و به‌طور کلی به فرآیند کشف دانش از داده‌ها اشاره می‌کند؛ در حال‌که داده‌کاوی به مرحله‌ای خاص از این فرآیند اشاره دارد». (Tuffery, 2002:)

محمدرضا رمضان‌پور و ریحانه خورسند (۱۳۹۷) با ترجمه کتاب *اصول کلان‌داده* اثر توماس ارل و پاول بوهرل^۱، ضمن بررسی مفاهیم، محرک‌ها و تکنیک‌های متن‌کاوی، به دلایل بهره‌مندی از این فناوری در دنیای مدرن اشاره کرده‌اند. حامد آقاپارزاده (۱۳۹۹) با ترجمه کتاب پیش‌بینی به وسیله متن‌کاوی اثر شلم م. وایس و تانگ ژانگ^۲، به کاربرد آن برای ارتقای علوم مختلف پیش از انجام پژوهش اشاره کرده است.

بهاره پهلوان‌زاده (۱۴۰۰) با ترجمه کتاب *تجزیه و تحلیل داده‌ها در ابر* اثر دومینکو تالیا و پائولو ترانیو^۳ کاربرد داده‌کاوی را به‌بوته نقد گذاشته و با نگاهی تازه به این فناوری نگرسته و مزایا و معایب آن را برشمرده است.

گفتنی است تاکنون پژوهشی مستقل مبتنی بر ضرورت بهره‌مندی از فناوری‌های نوین نظیر متن‌کاوی و داده‌کاوی^۴ در راستای مستندسازی و تکمیل مصاحبه‌های تاریخ شفاهی منتشر نشده است. بی‌تردید آشنایی با این تکنیک‌ها و کاربرد صحیح آنها، روشنایی‌بخش راه علاقه‌مندان و پژوهشگران تاریخ شفاهی خواهد بود.

کشف دانش و ارتباط با متن کاوی

امروزه با گسترش سیستم‌های اطلاعاتی، بهره‌مندی از ابزارهایی برای پردازش داده‌های ذخیره‌شده ضروری است تا کاربران از خلال این فرآیندهای نوین، به روابط منطقی میان اطلاعات پی ببرند. داده‌کاوی و زیرمجموعه‌های آن، از مهم‌ترین روش‌هایی هستند که از طریق آنها الگوهای مفید در داده‌ها با حداقل دخالت کاربران شناخته و تحلیل می‌شوند. درواقع با کاربرد تحلیل اکتشافی داده‌های مبتنی بر علم آمار، بر کشف اطلاعات ناشناخته تأکید می‌شود. (برومندزاده و چراغی‌فر، ۱۳۹۵: ۱۱)

برای ورود به بحث کاربرد متن‌کاوی در رفع ابهام^۵ و تکمیل منابع تاریخ شفاهی، یک مثال ساده در این زمینه ذکر

1. Thomas Erl & Paul Buhler.

2. Sholom M Weiss & Tong Zhang.

3. Domenico Talia & Paolo Trunfio.

4. Data Mining.

5. Disambiguation.

6. Web Mining.

7. Tuffery.

(1)

داده‌ها مواردی همچون پایگاه داده‌های ساخت‌یافته^۱، یا یک متن ساده غیرساخت‌یافته^۲ را شامل می‌شوند. گازندم^۳ چنین برداشتی از داده‌کاوی دارد: «داده‌کاوی به منظور شناخت الگوهای معتبر از داده‌های جمع‌آوری شده در مجموعه‌ای از پایگاه داده‌ها به کار می‌رود. کشف دانش مراحل تکراری و تعاملی^۴ را شامل می‌شود که ممکن است نیاز به تصمیم‌گیری کاربر داشته باشد» (Gazendam, 2010: 49).

متن‌کاوی یکی از مهم‌ترین بخش‌های داده‌کاوی و در مقایسه با آن پیچیده‌تر است. متن‌کاوی کشف و استخراج اطلاعات و دانش نهان، از منابع متنی است. دارونه^۵ هدف از متن‌کاوی را چنین بیان کرده است: «هدف از متن‌کاوی، جمع‌آوری اطلاعات غیرساخت‌یافته، استخراج شاخص معنادار عددی از متن و در نتیجه، فهم‌پذیرکردن و در دسترس قرار دادن اطلاعات موجود در متن برای الگوریتم‌های متعدد داده‌کاوی است. رده‌بندی (دسته‌بندی^۶ و طبقه‌بندی نیز گفته می‌شود)، خوشه‌بندی^۷ و خلاصه‌سازی متون^۸ از جمله کاربردهای متن‌کاوی است» (Darrooneh, 2011: 2).

به عبارت دیگر، متن‌کاوی عبارت است از فرآیند تحلیل متن به منظور استخراج اطلاعات از حجم عظیمی از متون غیرساخت‌یافته. متن‌کاوی شامل بازیابی اطلاعات، تحلیل متن، دسته‌بندی و طبقه‌بندی داده‌ها، یادگیری ماشینی و داده‌کاوی است. عده‌ای از جمله سباستین^۹، داده‌کاوی را مترادفی برای کشف دانش در نظر می‌گیرند و چنین درکی از آن دارند: «داده‌کاوی شامل تمام جنبه‌های فرآیند کشف دانش از پایگاه داده‌ها است و در واقع همان انتخاب روش و الگوریتم برای جست‌وجوی الگو در داده‌ها قلمداد می‌شود» (Sebastian, 1999: 13).

وب‌کاوی پایه پردازش زبان‌های طبیعی را در صفحات وب تشکیل می‌دهد. وب‌کاوی از مکانیزم‌های موجود

متن‌کاوی کلاسیک استفاده می‌کند تا روی ساختارهای متنی عمل کند و نیز با مکانیزم‌های هوشمند پردازش، داده‌ها را در وب فراخوانی می‌کند. وب‌کاوی به صورت کلی در ترکیب با سایر روش‌های داده‌کاوی استفاده می‌شود؛ به طور مثال، وب‌کاوی در فعال‌سازی شخصی‌سازی خودکار بر پایه استفاده وی‌بی به کار می‌رود. (حسن دوست، ۱۳۹۶: ۲۶)

اصطلاحنامه کتابداری، کلیدواژه‌ها یا کلمات کلیدی را واژه‌هایی تعریف می‌کند که به طور خلاصه و دقیق موضوع یا جنبه‌ای از موضوع را که در یک سند از آن بحث به میان آمده است، شرح می‌دهند. هم‌تک‌واژه‌ها و هم‌عبارت، کلیدواژه تعریف می‌شوند. (سلطانی و راستین، ۱۳۶۵: ۴۸) پورتر^{۱۰} درباره مزایای انتخاب کلمات کلیدی چنین دیدگاهی دارد: «گاه ممکن است مجموعه‌ای از کلمات کلیدی با واژگان از پیش تعریف‌شده محدود شده باشند. هدف این است که مجموعه‌ای کوچک از اصطلاحات استخراج شوند که یک سند را توصیف کند. سادگی و سازگاری از جمله مزایای انتخاب کلیدواژه‌ها مبتنی بر این روش است» (Porter, 1980: 130).

تکنیک‌های متن‌کاوی

همان‌طور که پیش‌تر ذکر شد متن‌کاوی به منظور استخراج اطلاعات و دانش از منابع مختلف غیرساخت‌یافته استفاده می‌شود. توجه به تکنیک‌های زیر در فرآیند متن‌کاوی ضروری است:

۱. «پیش‌پردازش متن^{۱۱}: به پاکسازی اسناد پیش از متن‌کاوی، پیش‌پردازش گفته می‌شود. لازم است داده‌هایمان از وجود عوامل نامناسب^{۱۲} پاک و ویژگی‌های اضافی آن حذف شوند. تیمورپور و نجفی پیش‌پردازش متن را به سه مرحله تقسیم می‌کنند:

۱.۱. «تقسیم‌بندی^{۱۳}: فرآیند جداسازی و شکستن هر کلمه در سند براساس واحدهای با معنا؛

۱.۲. ضریب تأثیر برای عبارات به تفکیک بخش‌ها؛

1. Structured.
2. Unstructured.
3. Gazendam.
4. Iterative.
5. Darrooneh.
6. Classification.
7. Clustering.
8. Text Summarization.
9. Sebastian.

¹⁰. Porter.

¹¹. Preprocessing.

¹². Noise.

¹³. Classification.

بیشتری اطلاعات و موضوعاتی را که به آن علاقه‌مندند، بررسی کنند. (پارسا، ۱۳۹۳: ۴۷)

شیدایی روش‌های خلاصه‌سازی را از نظر فضای زبانی^۶، به دو گروه تقسیم می‌کند:

۳،۱،۱. «رویکردهای کم‌عمق^۷: این روش، محدود و منحصر به سطح نحوی^۸ است و قطعات برجسته را از متن استخراج می‌کند. در این روش، جمله‌ها و پاراگراف‌های مهم انتخاب می‌شوند و یک نسخه کوتاه‌شده از متن را تشکیل می‌دهند. اهمیت جمله یا پاراگراف براساس ویژگی‌های آماری و گاهاً زبانی تعیین می‌شود.

۳،۱،۲. رویکردهای عمیق^۹: این روش شامل سطح معناشناسی^{۱۰} و پردازش زبانی^{۱۱} است. در این رویکرد مفهوم متن و درک آن با توجه به پردازش زبان‌های طبیعی بیان می‌شود». (شیدایی، ۱۳۹۱: ۳۸).

۳،۲. دسته‌بندی: این روش به‌طور عمده روی مجموعه‌ای از متون اعمال می‌شود و یک روش نظارت‌شده است و بدین منظور، از یک مدل آموزشی استفاده می‌شود. هدف اصلی این روش، طبقه‌بندی براساس نمونه‌های شناخته‌شده است و پس از آن، موارد ناشناخته به‌صورت خودکار طبقه‌بندی می‌شوند. این فرآیند، موضوع اصلی متن را با اضافه‌کردن داده‌های آموزشی و تجزیه و تحلیل محتوا روشن می‌سازد. در این تکنیک، پس از بازنمایی تعداد کلمات، درباره موضوع متن تصمیم‌گیری می‌شود. (خلیلی، ۱۳۹۱: ۳۶)

۳،۳. خوشه‌بندی: از این تکنیک برای تقسیم‌بندی متون مشابه استفاده می‌شود و متفاوت از دسته‌بندی است. در این روش، الگوهای ورودی و خروجی از پیش تعریف نشده‌اند. در خوشه‌بندی هیچ برچسب و کلاس از پیش تعریف شده‌ای وجود ندارد و به جای آن از مقدار شباهت بین موضوعات مختلف استفاده می‌شود. اسنادی که مشابه‌اند، در یک کلاس و مواردی که مشابه نیستند، در کلاس‌های متفاوت قرار می‌گیرند؛ بنابراین، هر خوشه شامل تعدادی متن است و موارد

۱،۳. فیلترکردن^۱: فرآیند حذف کلمات غیرمهم؛

۱،۴. ریشه‌یابی^۲؛

۱،۵. انتخاب مشخصه؛

۱،۶. وزن‌دهی به کلمات» (تیمورپور و نجفی، ۱۳۹۴: ۶۵).

۲. پس‌پردازش متن^۳: به معنای پاکسازی، هرس کردن، مرتب کردن و تکنیک‌های عمومی استفاده‌شده برای ارتقای نتایج به‌دست آمده از مراحل پیشین است که عملیات متن کاوی هسته‌ای^۴ نیز نامیده می‌شود. لوهن^۵ متن کاوی هسته‌ای را چنین توصیف می‌کند: «متن کاوی هسته‌ای شامل کشف الگوها و دانش، استخراج اطلاعات احتمالی و قطعی، تحلیل گرایش، شمارش تکرار ترم‌ها، و خوشه‌بندی است. مرور متن‌ها شامل فیلترها، پرسش‌ها و جست‌وجوها در ابزارهای مصورسازی مثل گراف‌ها است. مرور داده‌ها کمک می‌کند با محیط کار داده کاوی و پروسه‌های پشت آن ارتباط برقرار کنید». (Luhn, 2008: 160)

۳. برای استفاده از تکنیک‌های متن کاوی به‌منظور پردازش متن، باید الگوریتم‌های مناسبی انتخاب و اعمال شوند؛ الگوریتم‌هایی مانند خلاصه‌سازی، دسته‌بندی، خوشه‌بندی و اکاوی معنایی:

۳،۱. خلاصه‌سازی: اگرچه کامپیوترها قادرند افراد، مکان‌ها و زمان‌ها را شناسایی کنند، تجزیه و تحلیل معناشناسی همچنان دشوار است. زمانی که متن خلاصه می‌شود، باید با خواندن آن به یک درک کامل از آن برسیم. در نظر گرفتن نکات اصلی در نوشتن خلاصه امری ضروری است. به دلیل اینکه رایانه‌ها هنوز قابلیت درک زبان انسان را ندارند، باید روش‌های جایگزینی در پیش گرفت. یکی از این راهبردها، استخراج جمله‌های مهم است. همچنین، یکی دیگر از ابزارهای خلاصه‌سازی به‌منظور شناسایی نقاط کلیدی متن، جست‌وجوی عنوان یا سایر نشانه‌های فرعی است. بسیاری از ابزارهای خلاصه‌سازی متن به کاربر اجازه حق انتخاب متن را می‌دهند. افراد با خلاصه‌سازی قادر خواهند بود با سرعت

6. Linguistic space.

7. Shallow approaches.

8. Syntactic.

9. Deeper approaches.

10. Semantic level.

11. Linguistic Processing.

1. Filtering.

2. Rooting.

3. Postprocessing.

4. Core text mining operation.

5. Luhn.

عبارت سؤال، استخراج و پس از آن کلماتی که با آن رابطه معنایی دارند، جست‌وجو می‌شوند؛ در نتیجه، پس از کشف کلیه اطلاعات درباره آن موضوع، با توجه به سؤال مطرح شده، پاسخ مناسب داده می‌شود. همچنین، اساس شبکه واژگانی^{۱۰} بر ترادف‌ها و روابط معنایی میان کلمات حوزه‌ای است که این شبکه برای آن ساخته شده است. ایجاد و توسعه خودکار شبکه واژگانی به شدت به استخراج روابط معنایی وابسته است. این امر با توجه به این فرض انجام می‌گیرد که کلماتی که رابطه عام - خاص یا مترادف دارند، ویژگی‌های مشابهی نیز دارند؛ در نتیجه، با دانستن ویژگی‌های یک کلمه، ویژگی‌های کلماتی که با آن، رابطه معنایی خاص دارند، آشکار می‌شود. روابط معنایی استخراج شده، زیربنای بسیاری از سیستم‌های تولید و نمایش دانش الکترونیکی مانند انواع آنتولوژی‌ها^{۱۱}، سیستم‌های رده‌بندی^{۱۲} و سیستم‌های خبره^{۱۳} را تشکیل می‌دهد. (Clark, 2013: 65)

فیلدمن^{۱۴}، تلاش‌های انجام شده در زمینه استخراج روابط معنایی را به چهار دسته کلی تقسیم می‌کند:

- «بر اساس پردازش‌هایی که در حوزه روش‌های مبتنی بر الگو انجام می‌شود، معمولاً با توجه به الگوهای لغوی - نحوی موجود در متن، کلماتی که در قالب این الگو قرار می‌گیرند، به عنوان رابطه معنایی استخراج می‌شوند. یکی از مزایای مهم روش‌های بر پایه الگو نسبت به سایر روش‌ها این است که روابط معنایی به دست آمده از این روش‌ها روابطی با نام بوده‌اند و احتیاج به کار دستی برای تشخیص نوع رابطه نیست.

- روش‌هایی که اساس آنها بر تجزیه کردن متن ورودی و تبدیل آن به درخت تجزیه است. کارایی این روش به تجزیه‌گری^{۱۵} بستگی دارد که از آن استفاده می‌شود. اگر تجزیه‌گر استفاده شده مناسب باشد، درخت تجزیه به دست آمده،

مشابه در یک خوشه قرار می‌گیرند. (قهرمانی، ۱۳۹۶: ۵۳)

۳،۴ واکاوی معنایی: یکی از بخش‌های رایج در واکاوی اطلاعات، استخراج روابط معنایی^۱ میان عبارات موجود در متن است. روابط معنایی به رابطه موجود بین دو مفهوم بیان شده از یک کلمه یا عبارت اطلاق می‌شود. این روابط نقش بسیار اساسی در معناشناسی واژگان دارند. روابط معنایی متنوعی در حیطه زبان‌های طبیعی تعریف‌پذیر است که رابطه عام^۲ - خاص^۳، کل^۴ - جزء^۵، تضاد^۶ و ترادف^۷ معنایی از رایج‌ترین این روابط است. (خلیلی، ۱۳۹۱: ۳۴)

خاصه، کاربرد استخراج روابط معنایی در ایجاد پردازش‌های ماشینی روی متون را به صورت زیر شرح داده است:

۴،۳،۱ «خلاصه‌سازی متون: کشف شباهت بین متون؛

۴،۳،۲ پاسخگویی خودکار به سؤالات^۸: توسعه خودکار شبکه واژگانی؛

۴،۳،۳ استخراج اطلاعات از متون: تجزیه و تحلیل داده‌ها. (خاصه، ۱۳۸۹: ۸)

شایان ذکر است برای کشف شباهت میان متون، فرض بر این است که اگر در یک متن به جای یک کلمه، کلمات با رابطه عام - خاص یا مترادف قرار داده شوند، معنای آن جمله تغییر نمی‌کند. از این فرض بدین صورت استفاده می‌شود که برای محاسبه میزان شباهت بین دو متن، کلمات یکی از متن‌ها با کلماتی که رابطه عام - خاص یا مترادف با آن کلمه دارند، جابه‌جا می‌شوند. سپس تعداد کلمات مشابه دو متن شمرده می‌شوند و اگر از یک تعداد مشخص بیشتر شود، آن دو متن مشابه در نظر گرفته می‌شوند.

کلارک^۹ کشف شباهت میان متون بر اساس روابط معنایی را چنین توضیح داده است: «ابتدا کلمات کلیدی موجود در

1. Semantic relations.

2. Hypernym.

3. Hyponym.

4. Holonym.

5. Meronym.

6. Antonym.

7. Synonym.

8. Question Answering.

9. Clark.

10. Word net.

11. هستی‌شناسی

12. Taxonomy.

13. Expert systems.

14. Feldman.

15. Parser.

که در عین سادگی، دارای پیچیدگی و ظرافت‌های بسیاری است. مصاحبه تاریخ شفاهی معمولاً از دو شرکت‌کننده تشکیل می‌شود: مصاحبه‌کننده که در جایگاه مورخ قرار دارد و باید با روش پژوهش تاریخی آشنایی داشته باشد. مصاحبه‌شونده تاریخ شفاهی هم کسی است که در زمانه خود منشاء اثری یا دارای مشاهده و تجربه‌ای خاص باشد. در طرح‌های تاریخ شفاهی، مصاحبه‌شونده، منبع دست اول در مطالعات تاریخی محسوب می‌شود؛ زیرا او فعال در رویداد یا شاهد عینی رویدادی تاریخی یا مقطعی از تاریخ بوده است. (صالحی، ۱۳۹۹: ۳۱)

بنابراین، تعامل سازنده بین مصاحبه‌گر و راوی، در کیفیت مصاحبه تأثیر می‌گذارد و بخشی از فرآیند تاریخ شفاهی محسوب می‌شود. برخلاف یادداشت‌های شخصی که فرد به میل خویش ثبت می‌کند، در مصاحبه تاریخ شفاهی، راوی با همکاری یک مصاحبه‌گر آماده و مطلع، به تبادل اطلاعات در زمینه موضوع مصاحبه می‌پردازد که در نهایت، به درک تازه‌ای از آن منتهی می‌شود. مصاحبه‌گر ضمن هدایت بحث و با طرح سؤالات به‌جا و متقصدی، از ابهام مطالبی که راوی با الفاظ و عبارات خود بیان کرده است، می‌کاهد. کوینلن و دبلیوسامر، تمایز تاریخ شفاهی با گونه‌های مشابه آن را چنین تفسیر کرده‌اند: «تاریخ شفاهی بیانات برگرفته از دانش راوی نیست؛ بلکه مدرکی است که ساختار سؤالات پرسیده‌شده و پاسخ‌های آن را منعکس می‌کند؛ به همین دلیل، مستندکردن تمام مراحل تولید و پردازش مصاحبه، اهمیت متمایزی دارد تا به درک آیندگان از وضعیت انجام این ارتباط کمک کند». (کوینلن و دبلیوسامر، ۱۳۹۲: ۲۴)

مراحل فرآیند تاریخ شفاهی به‌طور خلاصه شامل این موارد است: انتخاب موضوع (زندگی یک شخصیت یا واقعه تاریخی یا موضوعی خاص یا ...)، تهیه طرح مصاحبه، پژوهش، تهیه سؤال، مستندسازی و راستی‌آزمایی در مرحله اول، ثبت سند مصاحبه (صوت، تصویر و ...)، تهیه نسخه‌های پشتیبان، پیاده‌سازی و ویرایش اولیه، تهیه فهرست توصیفی و تحلیلی، تدوین و ویرایش، فصل‌بندی، تعیین چارچوب،

مناسب خواهد بود و نتایج بهتری به دست می‌آید؛ در غیر این صورت، نمی‌توان به نتایج این روش‌ها اعتماد کرد.

• روش‌های آماری که روابط معنایی را براساس تعداد کلمات و نحوه قرارگرفتن آنها در کنار هم استخراج می‌کنند. این روش‌ها معمولاً رابطه‌های بی‌نامی را کشف می‌کنند و به کار دستی برای تشخیص نوع رابطه معنایی نیاز دارند. مزیت این روش‌ها در این است که مستقل از زبان‌اند و به یک یا چند زبان خاص محدود نمی‌شوند.

• روش‌های بر پایه منبع که از یک منبع خاص برای استخراج روابط معنایی استفاده می‌کنند. برای بهره‌مندی از این روش، به فرهنگ لغت مراجعه می‌شود». (Feldman, 2007: 38)

تاریخ شفاهی

تاریخ شفاهی به‌عنوان یکی از روش‌های نوین گردآوری اطلاعات، راهی پُرفرازونشیب را برای تثبیت جایگاه خود به‌منزله یک ابزار تحقیقی برای مطالعات تاریخی طی کرده است. این روش پژوهشی ویژگی روایی دارد و رویدادها را براساس گفته‌های ناظران دوره‌های تاریخی بازگو می‌کند. تاریخ شفاهی فرآیندی دوسویه یا حتی چندسویه است که در چارچوب مشخصی انجام می‌گیرد. (کوینلن و دبلیوسامر، ۱۳۹۲: ۱۸)

علاوه بر این، باید گفت تاریخ شفاهی یک روش پژوهش تاریخی برای جمع‌آوری و نگه‌داری اطلاعات از حافظه افراد سهیم در رخدادهای تاریخی یا شاهدان عینی آنها در هر طبقه اجتماعی است. تاریخ شفاهی سندی از جنس صوت و تصویر است. این روش پژوهشی در جست‌وجوی اطلاعات مستند شفاهی و در پی ثبت هر گونه تجربه انسانی است که به مرور زمان و در پی بروز تغییرات سیاسی، اجتماعی و فرهنگی در خطر نابودی قرار دارد. تاریخ شفاهی به بازآفرینی تاریخ کمک می‌کند و مورخ را به صحنه رخدادهای واقعی وارد می‌کند و این امکان را به او می‌بخشد که با یاری گرفتن از حافظه افراد، تاریخ را بازآفریند. (نیک‌نفس، ۱۳۹۲: ۶)

انجام مصاحبه، اصلی‌ترین فعالیت در تاریخ شفاهی است

موجود نباشد و هدف از انجام مصاحبه، تولید منبعی جدید باشد. یک مصاحبه‌گر حرفه‌ای پس از پایان هر جلسه، فهرست مواردی را که به تکمیل و مستندسازی نیاز دارند، تهیه می‌کند و در جلسات بعدی به‌طور دقیق بررسی می‌کند. در ضمن می‌تواند به روایت‌های سایرین در باب موضوعاتی که مصاحبه‌شونده براساس دیدگاه شخصی‌اش بیان کرده است، به دیده نقد بنگرد و دیدگاه آنان را در طرح سؤالات مدنظر قرار دهد.

علاوه بر این، مصاحبه‌کننده باید با زیرکی، بر انگیزه‌های آشکار و پنهان راوی متمرکز شود؛ زیرا در اموری که راوی نقش خود را منفی ارزیابی می‌کند یا آنکه قصد دارد در نقش خود بزرگ‌نمایی کند، ممکن است حقایق و واقعیت‌ها را به گونه‌ای دیگر بیان کند. در چنین مواردی است که مصاحبه‌گر با توجه به اهداف راوی از ذکر این موضوعات، با بیان اسناد معتبر مغایر با محتوا، به‌صورت کاملاً محترمانه، نسبت به صحت مطالب تشکیک وارد می‌کند.

بر این نکته تأکید می‌شود که پژوهشگران تاریخ شفاهی باید به بررسی داده‌های تاریخی در زمان انجام مصاحبه توجهی ویژه داشته باشند؛ زیرا با گذشت زمان، احتمال جابه‌جایی بسیاری از موارد در ذهن راوی وجود دارد و ممکن است او خودآگاه یا ناخودآگاه، بخش‌هایی از وقایع را کم یا زیاد کند. در تعیین اعتبار مطالب بیان‌شده، هرچند بنا بر صداقت راوی گذاشته می‌شود، از وظایف پژوهشگر است که به‌صورت منطقی، با روش گویاسازی و راستی‌آزمایی، اعتبار محتوای مصاحبه را کنکاش کند. (صالحی و تجلی، ۱۳۹۸:

۳۲)

۲ زمان تدوین متن مصاحبه: مستندسازی در این مرحله، فرآیندی است که طی آن تدوینگر برای فهمیدن تر و گویاتر کردن متن، با استفاده از منابع معتبر، درباره اشخاص، مکان‌ها، اصطلاحات، رویدادها، کتاب‌ها، مقاله‌ها، طرح‌های پژوهشی، وقایع تاریخی و ... ذکر شده در مصاحبه، توضیحات بجا، مناسب، مختصر و مفیدی در پانوشت صفحات یا پی‌نوشت‌های انتهای فصول درج می‌کند. حجم پانوشت‌ها یا

مستندسازی و گویاسازی در مرحله دوم، کنترل ارجاعات درون‌متنی، و آماده‌سازی پیوست‌ها.

مستندسازی تاریخ شفاهی با بهره‌مندی از متن کاوی

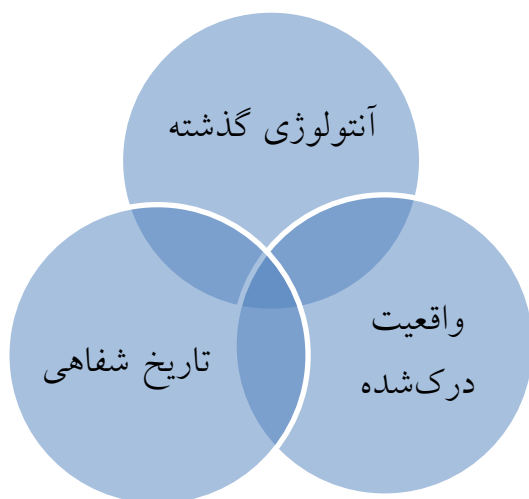
تاریخ شفاهی در دوران رشد و گسترش در قرن بیستم مراحل مختلفی را گذرانده است؛ البته در بیشتر موارد مراحل قبلی نیز به‌طور موازی با مرحله جدید به‌کار گرفته می‌شوند. دکتر مرتضی نورائی این مراحل را به شرح زیر برشمرده است:

- «در قرن بیستم که به اختراع ضبط‌صوت، با قابلیت حمل و دسترسی به شنیده‌ها منجر شد، ارجاع به مستندات آزمایش‌پذیر صورت پذیرفت؛
 - نیمه دوم قرن بیستم بر بازآفرینی و بازیابی و بازنگاری وقایع متمرکز بود؛
 - در دو دهه آخر قرن بیستم، سازوکار مرحله تحلیل و تفسیر وقایع در مدار ضبط تاریخ شفاهی‌نگاران قرار گرفت؛
 - در حال حاضر، دقت بر «حوالیات یا پیرامونیات» قرار دارد. درواقع مسائل حول و حوش وقایع در مرکز توجه مورخان تاریخ شفاهی جای گرفته است». (نورائی، ۱۳۹۹: ۴)
- در زمینه مستندسازی، راستی‌آزمایی و گویاسازی مصاحبه‌های تاریخ شفاهی باید به این نکته توجه داشت که مستندسازی در دو مرحله صورت می‌پذیرد:

۱ زمان انجام مصاحبه: مستندسازی در این مرحله، به مراتب مهم‌تر از انجام آن در زمان تدوین است؛ به دلیل اینکه ممکن است مصاحبه به مرحله تدوین نرسد؛ البته در صورتی که مصاحبه به مرحله تدوین برسد، باز این امکان وجود دارد که مصاحبه‌شونده در قید حیات نباشد تا بتوان به‌درستی به رفع ابهام‌ها و تکمیل مصاحبه اقدام کرد و به‌ناچار باید از منابع جنبی بهره برد.

مصاحبه‌گرانی که به موضوع مصاحبه و مسائل آن تسلط دارند، با طرح سؤالات تکمیلی، از مصاحبه‌شونده درخواست می‌کنند تا موارد گنگ بیان‌شده در مصاحبه را روشن‌تر بازگو کنند؛ به ویژه هنگامی که این اطلاعات در منابع مکتوب

نشود. آنچه از تاریخ انتظار می‌رود مستندسازی گذشته است؛ آنچنان که بوده یا دست‌کم آنچنان که درک شده است. هر قدر تاریخ شفاهی مستندشده بر واقعیت درک‌شده منطبق شود، پروژه گردآوری اطلاعات از روزن تاریخ شفاهی، اعتبار بیشتری پیدا می‌کند؛ همان‌طور که هر قدر واقعیت درک‌شده بر حقیقت منطبق شود، تاریخ‌نگاری ملموسی را براساس این روش پژوهشی در اختیار آیندگان قرار می‌دهد (نمودار ۱).



نمودار ۱: فضای مفهومی اعتبار منابع تاریخ شفاهی

فرآیند رفع ابهام از روزن متن‌کاوی

در سال‌های اخیر انجام مصاحبه‌های تاریخ شفاهی، بیشتر به‌منظور تولید داده‌های جدید و منحصربه‌فرد، معمولاً براساس دو روش زیر انجام می‌پذیرد (نمودار ۲):

- گردآوری و مکتوب‌سازی: در این روش، روایت، آنچنان که راوی به آن تمایل دارد، ثبت و ضبط می‌شود. در این حالت ممکن است نقش مصاحبه‌گر به حد اقل برسد و به یادآوری موارد دلخواه مصاحبه‌شونده بسنده شود؛ بدین صورت، مصاحبه به اهداف از پیش تعیین‌شده نمی‌رسد.

- مستندسازی و راستی‌آزمایی: راهکار علمی برای درک حقیقت (آنتولوژی) گذشته، اجرای طرح‌های شخص‌محور یا موضوع‌محور تاریخ شفاهی، مشروط بر این است که مستندسازی و راستی‌آزمایی جوهر غالب این مصاحبه‌ها باشد. لازم است در وهله نخست، به مستندسازی و سپس به راستی‌آزمایی توجه شود. مصاحبه‌کننده باید سؤالاتی را

پی‌نوشت‌ها هم به تناسب و به ضرورت و نیاز، کم و زیاد می‌شود.

لازم است تدوینگر در این مرحله با مصاحبه‌شونده - در صورت در قید حیات بودن او - تعامل سازنده‌ای داشته باشد و موارد مبهم را با او در میان بگذارد. همچنین، باید به موضوعاتی توجه کند که به توضیحات شفاف‌تر و مبسوط‌تری نیاز دارند. گویاسازی متن نشان می‌دهد تدوینگر بر موضوع تسلط دارد و برای مستندسازی و تبدیل مصاحبه به یک اثر پژوهشی تلاش کرده است. ناگفته نماند در این زمینه، سیاست سازمان مجری طرح تاریخ شفاهی نیز تعیین‌کننده است؛ با این حال، توصیه می‌شود جانب اعتدال حفظ شود و در این امر زیاده‌روی نشود. مستندسازی‌های مناسب در این مرحله، چهره‌ای نکته‌سنج از تدوینگر ارائه می‌کند. (مرادی‌نیا، ۱۳۹۷: ۱۳)

شماری از اهداف تدوینگر در مستندسازی و راستی‌آزمایی

مصاحبه‌های تاریخ شفاهی به شرح زیرند:

- رفع ابهام و معرفی اماکن، اشخاص، رویدادها، مفاهیم و به‌طور کلی هر واژه و مطلبی که به نظر او مبهم است؛
- توضیح بیشتر درباره آنچه مصاحبه‌شونده در مصاحبه بیان کرده است؛

- تکمیل مطالبی که راوی، آنها را درست و کامل توضیح نداده است؛

- توضیح درباره برخی واژه‌ها که راوی، درک آنها را از سوی خودش بدیهی می‌دانسته، ولی به نظر تدوینگر برای مخاطبان آن گنگ است؛

- شرح، تفسیر و تحقیق درباره برخی از مطالب متن؛

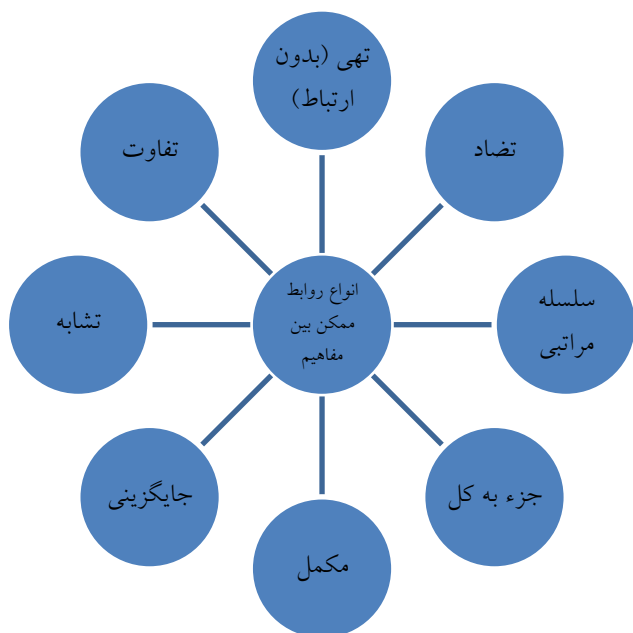
- رفع اشکالات محتوایی متن. (عبدالله‌زاده، ۱۳۹۵)

علاوه بر درج ضرورت مستندسازی مصاحبه‌ها در زمان انجام مصاحبه و نیز هنگام تدوین متن، باید به این موضوع توجه داشت که ممکن است محتوای برآمده از مصاحبه‌های تاریخ شفاهی در همه جنبه‌های آن، بدون انطباق با واقعیت درک‌شده و دوری از حقیقت (آنتولوژی) گذشته، اقدامی آگاهی‌بخش نبوده باشد و موجب گردآوری داده‌های تاریخی

جدول ۵: فهرست اسناد مکتوب مرتبط با محدوده زمانی، مکان‌ها، اشخاص و رویدادها.

پس از تکمیل اطلاعات جدول‌های مذکور، با بهره‌مندی از روش‌های متن‌کاوی که در سطور بالا شرح داده شد، اطلاعات مرتبط با کلمات کلیدی، از فهرست‌ها استخراج می‌شوند. سپس با تجزیه و تحلیل موارد فوق، ماتریس ارتباطات بین ابعاد مختلف موضوع اعم از مکان‌ها، محدوده زمانی، اشخاص، رویدادها و اسناد مکتوب تدوین می‌شود. باید تمام اجزای این ماتریس چندبعدی با هم روابط مشخص و بدون ابهامی را نمایش دهند.

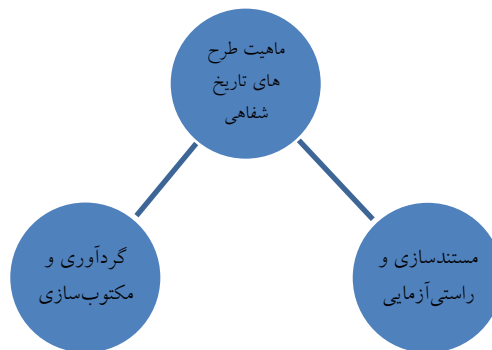
فهرست ابهام‌ها براساس انواع روابط ممکن بین مفاهیم (نمودار ۳)، زمینه طرح پرسش‌هایی را فراهم می‌آورد که لازم است از مصاحبه‌شونده پرسیده شود. متن کاوی ما را قادر می‌سازد تا ضمن تشخیص روابط پنهان بین مفاهیم، مستندات اکتشافی را با پاسخ‌های مصاحبه‌شونده، مقایسه و نتایج را به‌صورت تحلیلی، ثبت و نگه‌داری کنیم. همچنین، باید تضادها، اختلاف‌ها، مشابهت‌ها و هم‌جهتی‌ها را در مقایسه با مطالب برگرفته از مصاحبه با سایر اشخاص، روزآمد کنیم.



نمودار ۳: انواع روابط ممکن بین مفاهیم

۳. تحویل مستندات گام دوم به مصاحبه‌شونده و تکرار اقدامات انجام‌شده در گام دوم، براساس نتایج حاصله و انجام مصاحبه‌های تکمیلی.

براساس مطالب مندرج در کتاب‌ها یا گفته‌ها و مصاحبه‌های سایرین، طرح یا در گفته‌های مصاحبه‌شونده براساس سایر منابع تشکیک وارد کند. در این صورت ممکن است داده‌های گردآوری‌شده، از محتوای دلخواه مصاحبه‌شونده دور شوند؛ ولی به تاریخ‌نگاری شفاهی مبتنی بر مستندات تاریخی نزدیک تر می‌شود.



نمودار ۲: ماهیت طرح‌های تاریخ شفاهی

بنابراین، زمانی که ماهیت مصاحبه بر مستندسازی و راستی‌آزمایی در زمان انجام مصاحبه استوار باشد، فرآیند تاریخ شفاهی به سه مرحله یا گام (نمودار ۴) تقسیم می‌شود: ۱. مصاحبه‌شونده روایت را آن‌طور بیان کند که به آن علاقه‌مند است. ممکن است این گام چند جلسه از مصاحبه یا کل جلسات مصاحبه را به خود اختصاص دهد.

۲. تحویل مستندات گام اول به مصاحبه‌شونده و تکمیل پنجم جدول به شرح زیر:

جدول ۱: فهرست محدوده زمانی براساس اشخاص، رویدادها و مکان‌ها؛

جدول ۲: فهرست مکان‌ها براساس محدوده زمانی، اشخاص و رویدادها؛

جدول ۳: فهرست اشخاص براساس محدوده زمانی، مکان‌ها و رویدادها؛

جدول ۴: فهرست رویدادها براساس محدوده زمانی، مکان‌ها و اشخاص؛

مصاحبه توجه نشود، باید تدوینگر مدنظر قرار دهد.

مطالعه موردی: واقعه هفتم تیر ۱۳۶۰

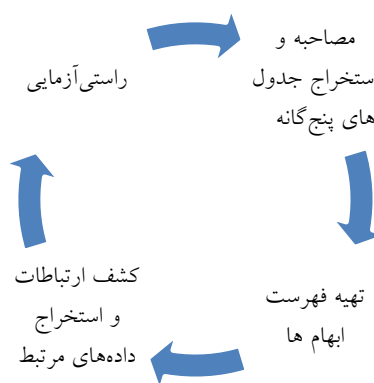
پس از بررسی روش‌های کشف دانش، تکنیک‌های متن کاوی و کاربرد آن به منظور مستندسازی و تکمیل مصاحبه‌های تاریخ شفاهی، موضوع «واقعه هفتم تیر ۱۳۶۰»، به عنوان مطالعه موردی بررسی می‌شود. در ارتباط با این موضوع، کتاب‌ها و مقاله‌های متعددی به رشته تحریر درآمده‌اند؛ از آن جمله: *تاریخ شفاهی شهدای هفتم تیر ۱۳۶۰*^۱، *روایت انقلاب و انفجار*^۲، *بهشتی با ۷۲ بال پرواز کرد*^۳، و *پرنده‌ای که تبعید شد*^۴.

گفتنی است با مطالعه مجموعه آثار مرتبط با موضوع مذکور و بهره‌مندی از ابزار متن کاوی در استخراج رویدادها، افراد، مکان‌ها و ... اطلاعاتی استخراج می‌شوند که به کشف روابط میان داده‌ها منجر می‌شود. روابط مکشوفه مبتنی بر فناوری‌های نوین موجب آشکارشدن ابهام‌هایی از این واقعه می‌شود که تا کنون پاسخ دقیق و روشنی به آن داده نشده است. از جمله این ابهام‌ها، ارتباط محمدجواد قدیری با واقعه هفتم تیر و دلایل شرکت نکردن بعضی از اعضای حزب جمهوری در آن جلسه هستند.

بنابراین، همان‌طور که مشاهده می‌شود اگر از ابزار متن کاوی در زمان انجام مصاحبه و در صورت نیاز در زمان تدوین متن، استفاده و کلیدواژه‌ها و موضوعات فرعی به دست آمده تجزیه و تحلیل شوند، موجبات دقت و سهولت در مستندسازی و جامعیت اطلاعات گردآوری شده را فراهم می‌آورد. در ادامه، برخی از کلمات کلیدی استخراج شده مربوط به رویدادها، مکان‌ها و افراد مرتبط با واقعه هفتم تیر ۱۳۶۰ (جدول ۱) درج می‌شوند. در ضمن تعدادی از ابهام‌های مکشوفه در ارتباط با این موضوع (جدول ۲) نیز ذکر

درخور ذکر است احتمالاً این سه مرحله بر حسب نیاز (نمودار ۴) بارها تکرار می‌شوند؛ تا جایی که مقدور است نسبت به رفع ابهام‌ها و تکمیل مصاحبه اقدامات لازم معمول شود. در ضمن اگر به مستندسازی در زمان انجام مصاحبه توجه نشود، تدوینگر می‌تواند با تکمیل همین جدول‌ها از مصاحبه‌شونده کمک بخواهد و نیز براساس سایر منابع برای رفع ابهام و تکمیل محتوا اقدام کند؛ بنابراین، تمرکز توأمان بر همه این موارد موجب مستندسازی و تکمیل مصاحبه در حین انجام مصاحبه و در صورت نیاز، در زمان تدوین متن می‌شود. از جمله اقدامات در این زمینه، موارد زیر ذکر می‌شوند:

- تهیه پانوشت‌ها یا پی‌نوشت‌ها و توضیحات و شواهد برای تبیین شباهت‌ها و تفاوت‌ها؛
- کنترل نهایی متن مصاحبه و پیوست‌های تکمیلی؛
- درج فهرست ابهام‌هایی که بی‌پاسخ مانده و لازم است به عنوان بخش مهمی از فرآیند تاریخ شفاهی در آینده به آن پاسخ داده شود.



نمودار ۴: چرخه پیشنهادی تاریخ شفاهی با رویکرد مستندسازی و راستی‌آزمایی

مصاحبه‌های تاریخ شفاهی پس از گذر از این مراحل کیفی، به منبعی با قابلیت استناد تبدیل می‌شوند. شایسته است به فرآیند مستندسازی در زمان انجام مصاحبه توجه شود و ابهام‌ها از راوی، پرسیده یا گفته‌های او به بوتۀ نقد گذاشته شوند؛ چون تمام مصاحبه‌ها با هدف تدوین و انتشار انجام نمی‌شوند؛ بلکه بسیاری از آنها به‌ویژه در آرشیوها و کتابخانه‌ها، تنها برای اطلاع‌رسانی و استفاده پژوهشگران ثبت و ضبط می‌شوند؛ البته اگر به این موضوع در زمان انجام

۱. صادقی، زهرا. (۱۳۹۵). *تاریخ شفاهی شهدای هفتم تیر ۱۳۶۰*. تهران: مرکز اسناد انقلاب اسلامی.

۲. فضلعلی، مرتضی. (۱۳۹۸). *روایت انقلاب و انفجار، مصاحبه با مرتضی فضلعلی*. مصاحبه محمدحسین یزدانی‌راد، تدوین و پژوهش پیمان صالحی. تهران: روزنه.

۳. بهشتی با ۷۲ بال پرواز کرد. (۱۳۶۳). تهران: ستاد برگزاری مراسم هفتم تیر.

۴. والایی، علی‌اکبر. (۱۳۸۵). *پرنده‌ای که تبعید شد: زندگینامه شهید علی‌اکبر سلیمی جهرمی*. تهران: شاهد.

می‌شود.

جدول ۲: تعدادی از ابهام‌های مکشوفه مرتبط با واقعه هفتم تیر ۱۳۶۰

ردیف	ابهام
۱	ارتباط محمدجواد قدیری با واقعه هفتم تیر
۲	دلایل شرکت نکردن بعضی از اعضای حزب جمهوری در آن جلسه
۳	دلایل نپذیرفتن مسئولیت حادثه توسط سازمان مجاهدین خلق
۴	نقش کشورهای خارجی در این واقعه
۵	چگونگی نفوذ در سازمان‌های امنیتی کشور
۶	نحوه ورود عاملان نفوذی در حزب جمهوری اسلامی
۷	نقش کشورهای خارجی در این واقعه
۸	نحوه دسترسی بمب‌گذار به بمب‌ها
و ...	

نتیجه

با توجه به مطالب درج شده در زمینه کاربرد ابزار متن‌کاوی و ذکر مطالعه‌های موردی به‌منظور مستندسازی و تکمیل مصاحبه‌های تاریخ شفاهی در زمان انجام مصاحبه و در صورت نیاز در زمان تدوین متن، نتایج زیر استنباط شدند:

- به دلیل گسترش اجرای طرح‌های تاریخ شفاهی، در بسیاری از موارد، بی‌توجهی به مستندسازی، راستی‌آزمایی و گویاسازی محتوای برآمده از این روش پژوهشی، این منابع را فاقد قابلیت استنادی کرده است. به دلیل اینکه این نقصان از زمان انجام مصاحبه پیش می‌آید و در صورت بی‌توجهی به آن تا هنگام تدوین ادامه پیدا می‌کند، لازم است با بهره‌مندی از روش‌های کارآمد، نسبت به رفع آن اقدام کرد.

- متن‌کاوی یکی از ابزارهای علمی و کاربردی برای تکمیل منابع تاریخ شفاهی است. استخراج روابط معنایی، بدین صورت که موضوع در مرکز قرار گیرد و مؤلفه‌های زمان، مکان، رویدادها و ... در حول آن به دقت تجزیه و تحلیل شوند، موجب به دست آمدن تاریخ‌نگاری شفاهی بر پایه مستندات می‌شود. تلاش به‌منظور مستندسازی و راستی‌آزمایی از سوی مصاحبه‌گر یا تدوینگر، هم موجب طرح پرسش‌های مجدد از مصاحبه‌شونده می‌شود و هم در حین تدوین، به انتشار متنی قابل استناد کمک شایان توجهی می‌کند. موارد زیر، از جمله مزایا و ارزش‌آفرینی‌های این رویکرد فناورانه‌اند:

- سرعت و سهولت در به‌کارگیری؛

جدول ۱: برخی از کلمات کلیدی استخراج‌شده مرتبط با واقعه هفتم

تیر ۱۳۶۰

رویدادها (به ترتیب زمان)	مکان‌ها	افراد (حقیقی و حقوقی)
عزل ابوالحسن بنی‌صدر	ساختمان حزب جمهوری اسلامی	شهادت: سید محمد حسینی بهشتی، رحمان استکی، سید رضا پاک‌نژاد، عباس حیدری، غلامحسین حقانی و ۶۸ شهید دیگر
تشکیل شورای موقت ریاست جمهوری	خیابان سرچشمه	جان‌بازان: علی اصغر باغانی، ایرج صفاتی دزفولی، مرتضی محمودی، مرتضی فضلعلی و ۲۴ جان‌باز دیگر
واقعه مسجد ابوذر	خیابان شهید مصطفی خمینی	محمد رضا کلاهی
رای‌گیری تغییر دستور جلسه	خیابان صیرفی‌پور	محمدجواد قدیری
بررسی ویژگی‌های رئیس جمهور	بیمارستان انقلاب	ساختار سازمانی حزب جمهوری اسلامی
سخنرانی آیت‌الله بهشتی	درمانگاه شهید شوریده	اعضای غایب حزب جمهوری اسلامی
سازمان مجاهدین خلق	بیمارستان طرفه	سپاه پاسداران انقلاب اسلامی
جلسات دفتر نخست‌وزیری	سازمان تبلیغات اسلامی	ابوالحسن بنی‌صدر
جلسات دفتر هماهنگی‌های رئیس جمهور	مسجد شهید مطهری	مسعود رجوی
ربودن مجروحان از بیمارستان‌ها	قطعه ۲۴ بهشت زهرا	امام خمینی (قدس سره)
اولین جلسه مجلس شورای اسلامی	آرامگاه شهدای هفتم تیر	آیت‌الله سید علی خامنه‌ای
انفجار دفتر نخست‌وزیری	موزه یادگار شهدای هفتم تیر	حجت‌الاسلام اکبر هاشمی رفسنجانی
انتخابات ریاست جمهوری	یادمان شهدای هفتم تیر	آیت‌الله حسین علی منتظری
شهادت امامان جمعه	بنیاد شهدای هفتم تیر	آیت‌الله سید عبدالکریم موسوی اردبیلی
و ...		

- <https://sarv1.blogfa.com>
- قهرمانی، عاطفه (۱۳۹۶). استخراج کلمات کلیدی با استفاده از تکنیک‌های متن کاوی. پایان‌نامه کارشناسی ارشد، دانشگاه گیلان، دانشکده فنی و مهندسی.
- کوینلن، مری‌کی؛ دبلیوسامر، باربارا (۱۳۹۲). راهنمای تاریخ شفاهی، ترجمه رضا مهاجر. تهران: سازمان اسناد و کتابخانه ملی ایران.
- مرادی‌نیا، محمدجواد (۱۳۹۷). گویاسازی معرف تسلط نویسنده بر متن است. نشریه الکترونیکی تاریخ شفاهی، ۷ (۳۵۲): ۱۳.
- نورائی، مرتضی (۱۳۹۹). گفت‌وگو با استاد دکتر مرتضی نورائی، مورخ تاریخ شفاهی و بحران کرونا. بازبینی‌شده در دوم مهر ۱۳۹۹، از: <http://t.me/mardomnameh>
- نیک‌نفس، شفیقه (۱۳۹۲). درآمدی بر تاریخ شفاهی. در شیوه‌نامه تاریخ شفاهی سازمان اسناد و کتابخانه ملی ایران. بازبینی‌شده در هفدهم تیر ۱۳۹۸، از <http://www.nlai.ir>
- Clark, Jonathan (2013). *Text mining and Scholarly publishin*. Research Consortium.
- Darooneh, Beta (2011). Keyword extraction by nonextensivity measure. *Alloation*, 50 (2): 2.
- Gazendam, Marta (2010). *Thesaurus Based Term Ranking for Keyword Extraction*. NewYork: Routledge.
- Luhn, Henry (2008). The automatic creation or literature abstracts. *IBM*, 2 (2): 159- 160.
- Meena, Kumar (2015). *Feature Based Sentence Filtering Method for Extractive Automatic Text Summarization*. India: Bubanswar.
- Porter, Sandi (1980). An algorithm for sux stripping. *Program*, 14 (3): 130.
- Sabestian, John (1999). Machine learning in Automate Text categorization. *ACN computing survey*, 34 (1): 13.
- Tuffery, Sebastiani (2002). Machine learning. In *Automated Text categorization*, 12 (1): 1.

- جامعیت در کشف روابط معنایی؛
- دقت در تجزیه و تحلیل موضوعات؛
- ایجاد بانک‌های اطلاعاتی با موضوعات مختلف؛
- رفع ابهام از محتوا و تکمیل فرآیند تاریخ شفاهی.

منابع

- برومندزاده، مصطفی؛ چراغی‌فر، سعید (۱۳۹۵). بررسی روش‌های متن کاوی. مشهد: مینوفر.
- پارسا، سودابه (۱۳۹۳). توسعه و بهبود روش‌های متن کاوی در کاربردهای فارسی. پایان‌نامه کارشناسی ارشد، دانشگاه شهید باهنر کرمان.
- تیمورپور، بابک؛ نجفی، حیدر (۱۳۹۴). داده کاوی با R به همراه متن کاوی و تحلیل شبکه‌های اجتماعی. تهران: مرکز تحقیقات و توسعه سازمان اتکا.
- حسن دوست، صالح (۱۳۹۶). خلاصه‌سازی متون با کاربرد روش‌های متن کاوی. پایان‌نامه کارشناسی ارشد، دانشگاه دیلمان لاهیجان.
- خاصه، علی‌اکبر (۱۳۸۹). داده کاوی، متن کاوی و وب کاوی: تعاریف و کاربردها. ماهنامه ارتباط علمی، ۱۶ (۲): ۸.
- خلیلی، مهدی (۱۳۹۱). استخراج روابط معنایی بین کلمات فارسی با استفاده از داده کاوی متن‌ها. پایان‌نامه کارشناسی ارشد، دانشگاه بوعلی سینا، دانشکده فنی و مهندسی.
- سلطانی، پوری؛ راستین، فروردین (۱۳۶۵). اصطلاحنامه کتابداری. تهران: کتابخانه ملی ایران.
- شیدایی، نوید (۱۳۹۱). متن کاوی متون فارسی در راستای پیش‌پردازش و دسته‌بندی مقالات خبری فارسی. پایان‌نامه کارشناسی ارشد، دانشگاه صنعتی اصفهان.
- صالحی، پیمان (۱۳۹۹). مسائل حقوقی مصاحبه‌های تاریخ شفاهی. تهران: سازمان اسناد و کتابخانه ملی ایران.
- صالحی، پیمان؛ تجلی، آرزو (۱۳۹۸). دستنامه ویرایش و تدوین آثار تاریخ شفاهی. تهران: معین ادارات.
- عبدالله‌زاده، محمد مهدی (۱۳۹۵). دیدگاه‌هایی درباره گویاسازی. بازبینی‌شده در بیستم تیر ۱۴۰۰، از

