

Eine graphbasierte Indexstruktur zum inhaltsbasierten Audioretrieval

Alexander Höck¹, Frank Kurth¹, Michael Clausen²

¹ Fraunhofer FKIE, 53343 Wachtberg, Deutschland, Email: {alexander.hoeck, frank.kurth}@fkie.fraunhofer.de

² Universität Bonn, 53113 Bonn, Deutschland, Email: clausen@informatik.uni-bonn.de

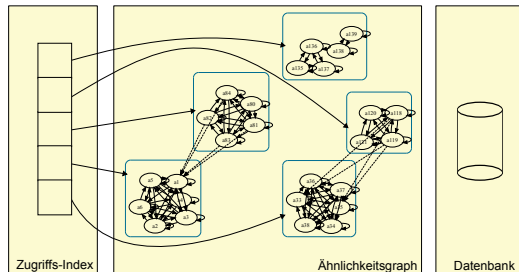


Abbildung 1: Schematische Struktur des graphbasierten Indexes, links: Zugriffs-Index, Mitte: Ähnlichkeitsgraph mit angelegter Clusterstruktur, rechts: zugrunde liegende Audio-datenbank.

Einleitung

Graphbasierte Retrieval- bzw. Ranking-Verfahren haben insbesondere im Bereich des Web-Retrievals zu großen Qualitätssteigerungen der Retrievalergebnisse geführt. Es liegt daher nahe, diese Verfahren in geeigneter Weise für allgemeine Aufgaben im Bereich Multimediaretrieval zu adaptieren. In diesem Beitrag stellen wir vor diesem Hintergrund eine neuartige Indexstruktur zum inhaltsbasierten Audiomatching vor, deren Kernkomponente ein automatisch generierter Ähnlichkeitsgraph für Merkmalsfolgen ist. Hierbei läuft die Suche nach zu einer Anfrage ähnlichen Abschnitten innerhalb einer großen Kollektion von Musikaufnahmen basierend auf dem in einem Vorverarbeitungsschritt aufgebauten Index in drei Schritten ab. Zunächst werden wenige bezüglich der Anfrage vielversprechende Knoten im Ähnlichkeitsgraphen identifiziert. Ausgehend von diesen Knoten werden im zweiten Schritt weitere Trefferkandidaten anhand der Struktur des Graphen ausgewählt. Abschließend werden die selektierten Kandidaten exakt bezüglich der Anfrage bewertet und ausgegeben. Die vorgestellte Indexstruktur wird mit bekannten Verfahren zum Audiomatching [1, 2] verglichen und mögliche Einsatzszenarien des graphbasierten Ansatzes werden vorgeschlagen.

Indexkonstruktion

Die graphbasierte Indexstruktur besteht im Wesentlichen aus zwei Komponenten (siehe Abbildung 1). Als Kernkomponente wird ein Graph verwendet, der in geeigneter Form Ähnlichkeiten zwischen kurzen Ausschnitten der Audiostücke widerspiegelt. Der Graph wird bei der Anfragebearbeitung zum Auffinden von Trefferkandidaten herangezogen. Um hierbei geeignete Startpunkte innerhalb des Graphen zu identifizieren, wird eine angepasste Indexstruktur [1] verwendet. Im Folgenden wird zunächst auf die Konstruktion des Graphen eingegangen und im

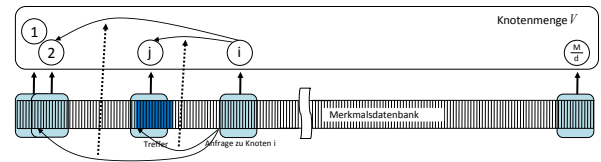


Abbildung 2: Skizze zur Konstruktion des Ähnlichkeitsgraphen; Zuordnung der Merkmalsblöcke zu Knoten-IDs sowie Berechnung der Kanten.

Anschluss auf die Erstellung des Zugriffs-Index. Vorweg sei erwähnt, dass für die graphbasierte Indexstruktur eine Vorverarbeitung der Audiodaten notwendig ist. Es werden die in [2] vorgeschlagenen CENS-Merkmale verwendet. Im weiteren Verlauf wird davon ausgegangen, dass die Merkmalsextraktion bereits stattgefunden hat und somit eine zur Audiodatenbank korrespondierende Merkmalsdatenbank existiert.

Der Ähnlichkeitsgraph soll Ähnlichkeiten zwischen Audioausschnitten der Datenbank widerspiegeln. Jeder Knoten korrespondiert zu einem gewissen Audioausschnitt und Kanten beschreiben die Ähnlichkeit inzidenter Knoten. Zur Konstruktion der Knoten wird die gesamte Merkmalsdatenbank in sich überlappende Abschnitte fester Breite L und fester Schrittweite d unterteilt. Der Beginn des Merkmalsbereichs von Knoten n ergibt sich implizit, $(n - 1) \cdot d + 1$ gibt das erste Merkmal des Bereichs an. Zur Berechnung der Ähnlichkeiten zwischen den Merkmalsbereichen der Knoten wird die indexbasierte Suche [1] verwendet. Für jeden Knoten wird hierzu der entsprechende Merkmalsbereich als Anfrage an einen auf der vollständigen Merkmalsdatenbank basierenden Index gestellt. Die auf diese Weise gefundenen ähnlichen Bereiche entsprechen in der Regel nicht exakt einem Knoten. Daher werden sie mittels Quantisierung dem nächstgelegenen Bereich zugeordnet (siehe Abbildung 2) und abschließend werden die gewonnenen Kanten dem Graphen hinzugefügt. Mit der Kantenkonstruktion ist die Grapherstellung abgeschlossen. Als Zugriffs-Index kommt erneut die Indexstruktur [1] zum Einsatz. Allerdings wird nicht die vollständige Datenbank indexiert, sondern lediglich ein repräsentativer Teil. Zur Selektion dieses Teils der Datenbank wird der zuvor konstruierte Graph analysiert. Es wird ein Clustering des Graphen berechnet und für jedes Cluster wird ein repräsentativer Merkmalsbereich der Datenbank ausgewählt. Die ausgewählten Merkmalsbereiche werden zu einem synthetischen Dokument D_R zusammengefasst. Hierbei werden angrenzende und sich überschneidende Merkmalsbereiche zusammengefasst, um Redundanzen zu vermeiden.

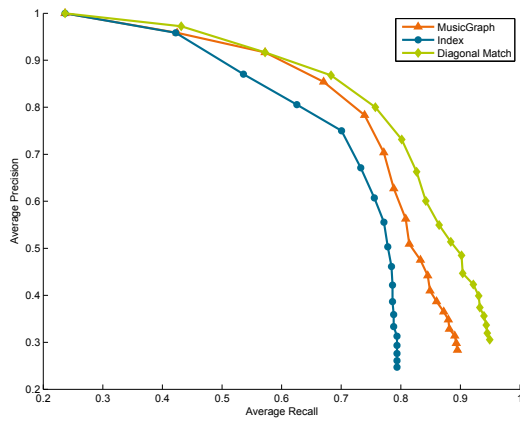


Abbildung 3: Average Precision-Recall-Diagramm über Ground-Truth Anfragen für indexbasierte Suche, graphbasierte Suche und Diagonal-Match.

Auf der Grundlage dieses synthetischen Dokuments wird dann mit den in [1] vorgestellten Verfahren der Zugriffs-Index erstellt.

Suche

Die graphbasierte Suche nach zu einer Anfrage ähnlichen Stellen innerhalb der indextierten Datenbank läuft in drei Schritten ab. Zunächst werden mittels des Zugriffs-Index geeignete Einstiegsknoten für den eigentlichen Suchschritt bestimmt. Im Anschluss werden durch Graphtraversierung weitere Trefferkandidaten identifiziert. Im letzten Schritt wird eine Trefferliste aus den Trefferkandidaten selektiert und ausgegeben. Bei der Suche nach Einstiegsknoten wird die Anfrage mittels indexbasierter Suche [1] auf D_R ausgewertet. Die Trefferstellen innerhalb D_R werden dann invers zur Konstruktion von D_R in die Menge der Einstiegsknoten V_E umgewandelt. Ausgehend von der Menge der Einstiegsknoten V_E werden im zweiten Suchschritt mit Hilfe einer angepassten Form der Breitensuche weitere Trefferkandidaten V_T identifiziert. Insbesondere werden hierbei Zusatzinformationen wie beispielsweise die Kantengewichtungen ausgewertet, die bei der Graphkonstruktion sowie der Suche nach Einstiegsknoten ermittelt wurden. Ziel der Anpassungen ist ein frühzeitiger Abbruch der Breitensuche. Abschließend werden alle gefundenen Trefferkandidaten V_T mittels Diagonal-Match [2] bezüglich der Anfrage bewertet. Alle Trefferkandidaten mit einer Bewertung größer einer Schranke werden als korrekte Treffer akzeptiert und in einer Trefferliste ausgegeben.

Evaluation und Ausblick

Das zuvor vorgestellte Verfahren zum inhaltsbasierten Audioretrieval wurde prototypisch in MATLAB implementiert und mit der indexbasierten Suche [1] und der Suche mit Diagonal-Match verglichen. Grundlage des Vergleichs war eine Audiodatenbank, die 112 Stunden klassischer Musik umfasst. Neben der Datenbank stand eine Sammlung von 36 Anfragen mit manuell annotierten Trefferlisten als Ground-Truth zur Verfügung. In Bezug auf die Antwortzeit bei der Anfragebearbeitung

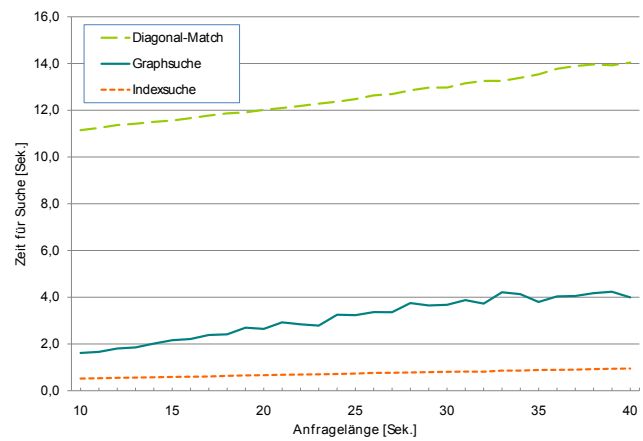


Abbildung 4: Durchschnittliche Zeit zur Bearbeitung einer Anfrage abhängig von Anfragelänge für indexbasierte Suche, graphbasierte Suche und Diagonal-Match.

stellt sich heraus, dass die Bearbeitungszeit einer einzelnen Anfrage im Mittel höher ist als bei der indexbasierten Suche. Der Unterschied liegt um etwa einen Faktor drei höher. Zwar kann gegenüber der indexbasierten Suche keine Verkürzung der Laufzeit erzielt werden, gegenüber der Suche mittels Diagonal-Match wird jedoch eine Verkürzung der Suchzeit von durchschnittlich 9,5 Sekunden erzielt, was einer Reduktion um circa 75% entspricht (siehe Abbildung 4). Es stellt sich ferner heraus, dass mit dem gewählten Verfahren qualitativ bessere Ergebnisse gefunden werden können, als mit der indexbasierten Methode [1]. Im Vergleich zur Suche mittels Diagonal-Match [2] liefert die graphbasierte Suche deutlich bessere Ergebnisse als die indexbasierte. Die graphbasierte Suche findet im Mittel die besten drei Trefferstellen übereinstimmend mit Diagonal-Match. Im Vergleich zur indexbasierten Suche hat sich eine Verbesserung von über 36% ergeben, für kurze Anfragen sogar bis zu 60%. Für einen Einsatz der graphbasierten Suche bieten sich verschiedene Möglichkeiten. Beispielsweise ist eine Kombination mit der indexbasierten Suche denkbar. Hierbei können dem Benutzer zunächst sehr effizient die Trefferstellen der indexbasierten Suche präsentiert werden. Im Anschluss können dem Benutzer weitere durch eine Nachverarbeitung mit der graphbasierten Methode gefundene Trefferstellen gezeigt werden. Ferner sind weitere Anpassungen der graphbasierten Suche mit speziellem Bezug auf die Laufzeit denkbar. Ansatzpunkte bieten hier vor allem das verwendete Clusteringverfahren bei der Indexerstellung sowie die Strategie zur Graphtraversierung während der Suche.

Literatur

- [1] Kurth F. und Müller M.: Efficient Index-Based Audio Matching. IEEE Trans. on ASLP, vol. 16, no. 2 (2008), 382–395
- [2] Müller M., Kurth F. und Clausen M.: Audio Matching via Chroma-based Statistical Features. In Proceedings of ISMIR, London, GB, (2005)