ANTS

— полная система для автоматической аннотации новостных программ на основе аудиовизуального контента и анализа текста

Giorgio Dimino, Alberto Messina и Roberto Borgotallo
RAI Centre for Research and Technology Innovation

В статье описывается интегрированная система для автоматической аннотации телевизионных новостных программ под названием ANTS (Automatic Newscast Transcription System). Она состоит из нескольких аналитических компонентов, интегрированных внутри единой архитектуры. Пользователи имеют возможность доступа к большой, ежедневно растущей базе новостных сюжетов с главных национальных каналов – идентифицированных, разбитых на категории и опубликованных с полной автоматизацией. Система идентифицирует границы сюжетов, извлекает текст из речевого контента, классифицирует сюжеты по темам и дает ссылки на внешнюю информацию из Сети.

Производительность системы была оценена в реальном сценарии группой профессиональных пользователей RAI. Сильная сторона подхода ANTS – способность к интеграции нескольких гетерогенных инструментов в высокопроизводительной и готовой к производству среде. ANTS может производить много часов материала в сутки без существенных перепадов и с достаточной точностью для промышленного внедрения в крупных вещательных компаниях.

Введение и сопутствующая работа

Автоматическая сегментация программ – одна из самых проблемных и сложных тем для исследования.

Хотя возможность корректной сегментации – ключевой фактор в повышении доступности и точности поиска и запросов, в решении этой проблемы в целом нельзя рассчитывать на установленный подход. Общая база принципов для новостей состоит из комбинации визуальных, звуковых и речевых характеристик.

Сегментация новостей входила в круг задач инициативы TRECVID [1] в 2003 и 2004 г. Труды, описанные в [2] и [3], иллюстрируют несколько разных подходов, определенных и разработанных участниками

TRECVID в этих двух сериях. Лучшие подходы, представленные в на TRECVID 2004, включали анализ видео и звука, отдельно или в сопровождении автоматического перевода речи в текст, и показали F-меру от 0.6 до 0.7. Базовые характеристики, использованные в нескольких случаях: (i) визуальное подобие между планами в рамках временного окна и (ii) временное расстояние между планами [4]. Другие эвристические правила – например, подобие лиц в планах и обнаружение повторов появления анкерных лиц [5][6][7] – могут добавить дополнительный уровень информации для повышения точности.

Контрибуция аудио канала может использоваться для обнаружения пауз, потенциальных границ смены тем [4][6][8] или для обнаружения изменений шаблонов классификации звука (например, с музыки на речь [6]) или смены диктора [8].

В качестве третьего источника информации очень часто используется текст из расшифровки или автоматического распознавания речи, либо поиском появления похожих слов в разных планах, либо обнаружением подобия текста между планами [5][6].

Использование автоматического перевода речи в текст вносит некоторые проблемы в задачу сегментации новостей из-за типичных ошибок типа пропуска, удаления и вставки слов, а также неверно расшифрованных слов. Современная система не использует данные Newsroom Computer System (NRCS), хотя мы недавно начали изучать, как интегрировать эту информацию, как для улучшения расшифровки, так и повышения производительности автоматической сегментации.

Архитектура

Система, главными компонентами которой являются централизованный модуль управления рабочими процессами и коллекция *AntsClients*, спроектирована сильно распределенной и масштабируемой.

Каждый *AntsClient* конфигурируется для выполнения определенной задачи общего процесса, например, перевода речи в текст и сегментации новостных сюжетов. Работая как демоны, *AntsClients* соединяются с модулем управления рабочими процессами через протокол HTTP для получения задач и уведомления об успехе /сбое. Затем запрашиваемый процесс выполняется определенной командой, выданной локально *AntsClients*. Такой подход, описанный на *Рис. 1*, предусматривает внедрение в нескольких хостах внутри сети и/или в нескольких экземплярах, поставляющих одну и ту же услугу, для возможности масштабирования системы по мере необходимости для достижения нужного результата и обеспечения средств для преодоления отказа.

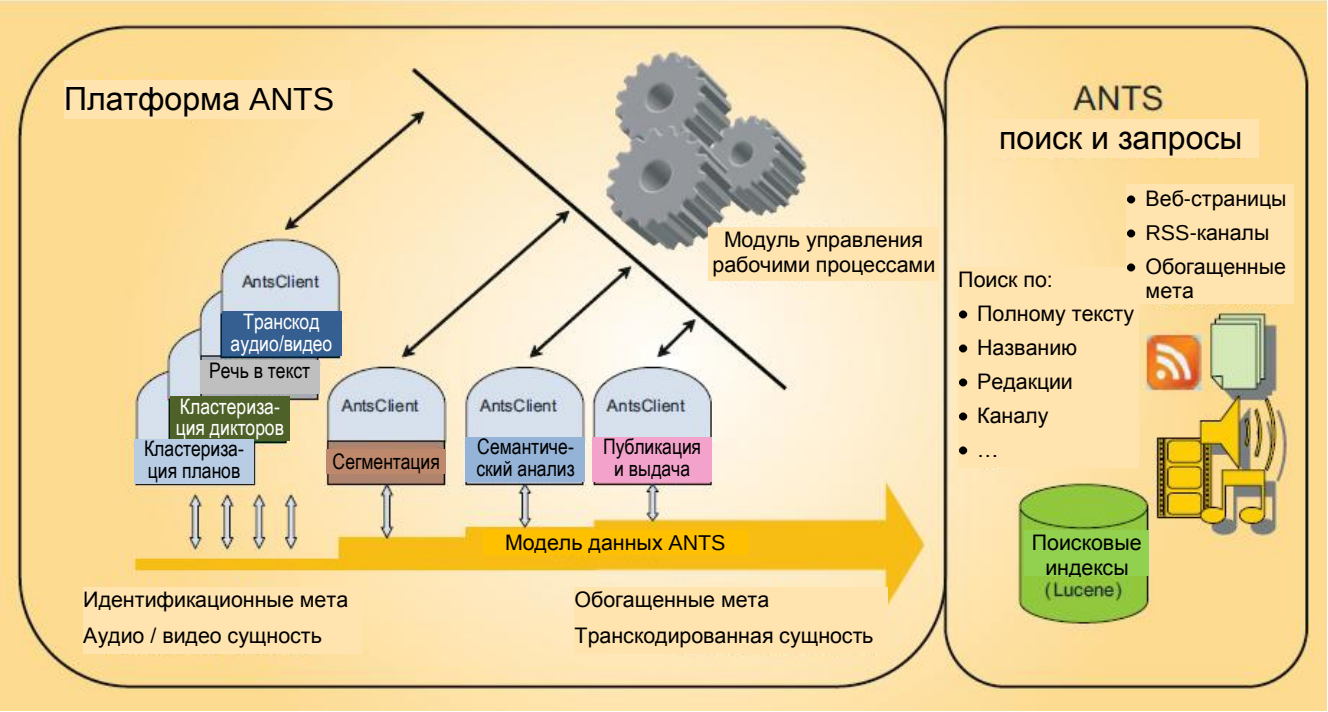

Рис. 1
Архитектура ANTS

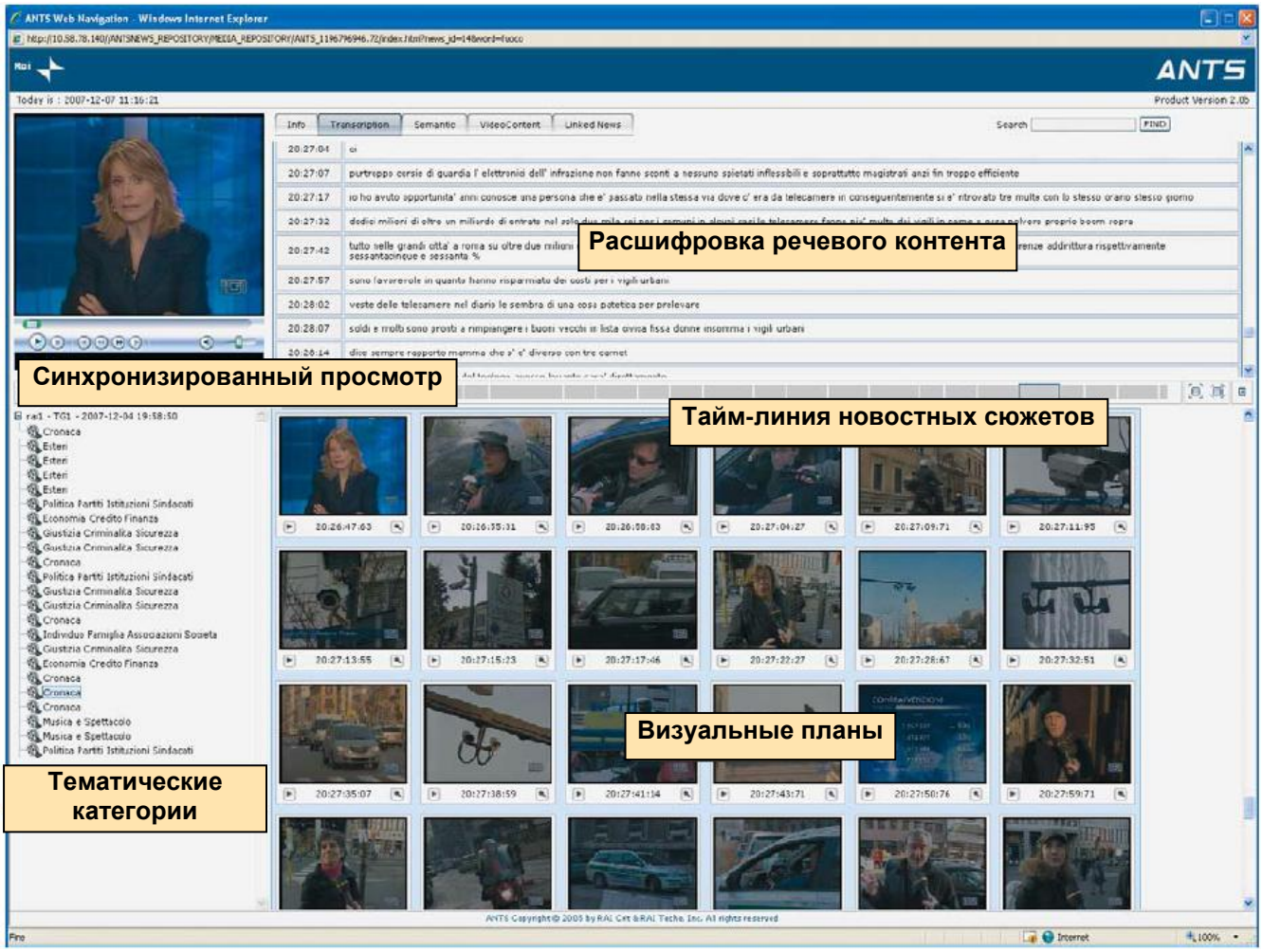

Рис. 2
Интерфейс браузера ANTS

Все произведенные метаданные собраны в централизованном хранилище до конечной передачи на платформу публикации. Подсистема поиска и запроса поддерживает поиск по полному тексту, с фильтрацией по категориям и/или именованным объектам, помимо идентификации и информации о публикации. Наконец, как показана на *Рис. 2*, представленная страница отображает все результирующие компоненты и позволяет их синхронно включение по общей тайм-линии. Все функции запроса – а также мониторинга и администрирования – доступны по IP сети посредством веб-браузера.

Кроме того, ANTS передает законченные материалы и метаданные, собранные в формате XML, в систему каталога долговременного архива RAI.

Инструменты анализа и службы автоматической аннотации

На *Рис. 3* показана функциональная блок-схема ANTS, включающая семантический анализ речевого текста и автоматическую редакторскую сегментацию.

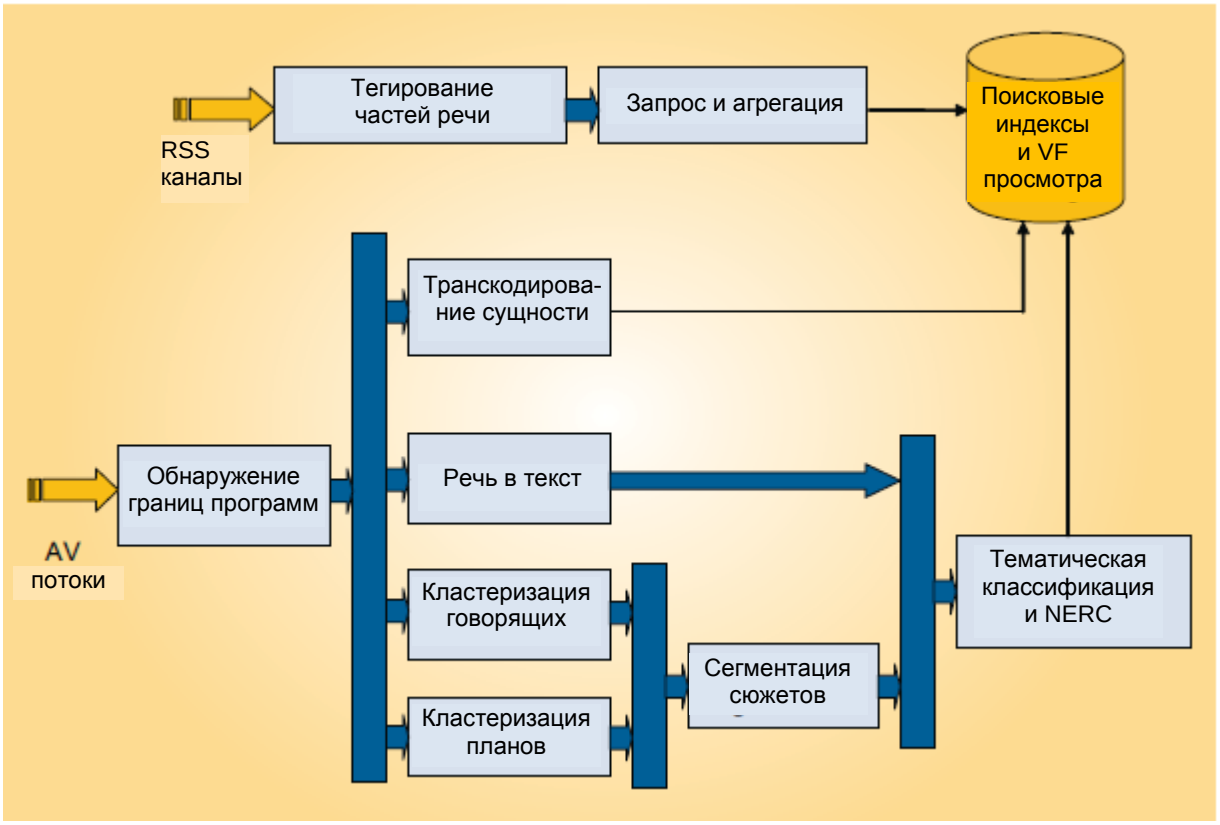

Рис. 3
Функциональная блок-схема ANTS

Распознавание видеоклипов

Для автоматической сегментации прямых потоков на программы, которые можно анализировать в последующей цепи, ANTS использует метод распознавания видеоклипов.

Начальные и конечные позывные программ используются как опора для поиска в полученных потоках посредством алгоритма кластеризации планов.

В автономном процессе из каждого плана опорных клипов извлекаются отличительные признаки, включая пространство цветов HSV[1] и гистограммы яркости, гистограммы текстуры (контрастность и направленность) и гистограммы временной активности.

Все гистограммы имеют 65 bins, образуя вектор признаков 65 x 7 x N для каждого плана, где N – число изображений в плане. Во время интерактивной фазы признаки планов используются как многомерные фиксированные центроиды в алгоритме кластеризации.

В конце кластеризации входящие планы, собранные с достаточной силой (т.е. близко к центроиду), классифицируются как экземпляры пана, связанные с центроидом. Средняя точность обнаружения этого процесса – 0.86, при возврате – около 0.87. (Точность определяется как соотношение между корректно идентифицированными элементами и общим числом обнаруженных элементов, а возврат – как соотношение между корректно идентифицированными элементами и эффективным числом элементов.)

Кластеризация планов

Кластеризация планов производится методом оптимизированной кластеризации снизу вверх, где в качестве измерения базового расстояния используется пересечение гистограмм признаков.

Весь процесс делится на следующие этапы:

[1] Объяснение HSV см. в Википедии: http://en.wikipedia.org/wiki/HSV_color_space

Сокращения

ANTS	(RAI) Automatic Newscast Transcription System (RAI) Система автоматической расшифровки новостей	**NRCS**	NewsRoom Computer System Компьютерная система ньюсрума
HTTP	HyperText Transfer Protocol Протокол передачи гипертекста	**RSS**	Really Simple Syndication Реально простая передача информации
IP	Internet Protocol Интернет-протокол	**XML**	eXtensible Markup Language Расширенный язык разметки

- извлечение гистограмм признаков;
- обнаружение планов;
- частичная кластеризация на сегменты постоянной длины;
- отбор соответствующих кластеров, найденных на этапе частичной кластеризации, и
- повторная кластеризация избранных кластеров.

Таким образом, процесс имеет на входе ряд изображений и набор кластерных ярлыков, связанных с каждым входящим изображением, на выходе. Это значит, что процесс производит обнаружение планов как побочный эффект процесса кластеризации.

Кластеризация звука

Кластеризация звука производится инструментом mClust [9], бесплатным программным обеспечением по публичной лицензии GNU, разработанным в лабораториях LIUM университета Мэна во Франции.

Результаты работы состоят из набора маркированных кластеров, каждый из которых указывает на человека, говорящего в проанализированном аудиоклипе. Один кластер – это набор временных интервалов с относительными границами от начала клипа.

Сегментация новостных сюжетов

Сегментация новостных программ на сюжеты производится посредством звуковых и визуальных cues с помощью трехуровневой эвристической структуры, выведенных путем наблюдения редакторских стилей статистически значимого набора программ, длительностью примерно 40 часов (~80 программ).

Базовая эвристика, широко принятая в литературе – например, в [10], – способна обнаруживать границы планов, содержащих анкерное лицо, эквивалентно обнаружению границ новостных сюжетов.

Для обнаружения планов с анкерным лицом мы применяем другую эвристику, а именно, что самый частый говорящий – это анкерное лицо и что он/она говорит периодами в течение всей программы. Это позволяет выбрать самого вероятного кандидата среди всех идентифицированных процессом кластеризации говорящих.

Этот подход не позволяет системе различать ситуации, в которых анкерное лицо вводит последовательно несколько коротких сюжетов без внешних материалов (например, репортажей). Для преодоления этого ограничения мы используем третью эвристику, состоящую в знании о том, что в подавляющем большинстве случаев введение нового краткого сюжета сопровождается сменой плана камеры (например, с крупного на дальний). Для оптимизации точности в выборе смены планов камеры мы осуществляем процесс кластеризации видеопланов на основе признаков, указанных на *Рис. 3*. Это позволяет обнаруживать и классифицировать кластеры планов как принадлежащие студийным съемкам с анкерным лицом, следуя той же эвристике частоты/расширения, что и в обнаружении диктора-кандидата. Этот процесс двойной кластеризации (звука и видео) дает очень простой и эффективный рекурсивный алгоритм, который альтернативно выбирает видео и аудио кластеры на основе их процента взаимного охвата.

Наконец, границы сюжетов идентифицируются там, где граница аудио или видео кластера находится среди кластеров, выбранных рекурсивным алгоритмом, с адаптивным порогом во избежание пересегментации. Схема экспериментальной оценки алгоритма сегментации представлена на следующей странице.

Тематическая классификация и ссылка на внешние источники

В ANTS извлечение речевого контента производится с помощью механизма перевода речи в текст на базе [11], способного расшифровывать итальянский и английский языки. Тематическая классификация сегментированных сюжетов производится по упрощенной байесовской модели классификации, обучен-

ной по фонду элементов извлеченного текста и аннотированной со стандартной тематической таксономией из 28 классов. Фонд насчитывает 25’000 элементов, 4/5 которых используются для обучения, а 1/5 для тестирования. Общая точность тематической классификации в работающей систем – 0.82, а точность программного уровня, т.е. средняя точность классификации, вычисленная по ряду элементов, принадлежащих одной программе – 0.88.

Ссылка на внешние источники информации реализуется через лингвистический анализ RSS-каналов шести крупных газет, где периодически проводятся опросы. По каждому названию элемента RSS производится тегирование частей речи для извлечения ключей самых значимых слов, которые в свою очередь используются для выполнения полнотекстового запроса по тексту, автоматически извлеченному из элементов теленовостей. Результаты запроса организуются в просмотровый список, связанный с названием элемента, и каждый отдельный сюжет связывается с элементом RSS, для которого поисковые баллы выше определенного порога. Агрегированные новостные сюжеты под определенным названием могут считаться RSS услугами, предоставляемыми ANTS. В результате пользователи получают мультимедийную интеграцию RSS-каналов из крупных газет с соответствующими новостными сюжетами, собранными из крупных информационных телепередач. С другой стороны, элементе теленовостей могут автоматически аннотироваться с помощью названий RSS.

Средняя точность процесса агрегации новостных элементов – 0.97, вычисленная как отношение между соответствующими новостными элементами и общим числом, относительно названия RSS-канала в определенной агрегации.

Экспериментальная оценка алгоритма сегментации новостных сюжетов

Мы протестировали наш алгоритм сегментации новостных сюжетов с рядом тестовых программ с материалом длительностью около 40 часов. Тестовый набор тегировался вручную, т.е. были идентифицированы все реальные границы сюжетов. Для оценки работоспособности системы мы использовали измерение с выравниванием с учетом начальных и конечных границ разного веса, а также с предположением, что недостающий материал больше влияет на измерение, чем излишний.

На первом этапе мы произвольно выбрали поднабор тестового материала и эмпирически оптимизировали параметры оценки для достижения совпадения между пользовательскими оценками качества сегментации и объективным измерением. Таким образом, мы получили измерение качества, подтвержденное пользователями. На втором этапе, после подтверждения измерений по описанной процедуре, мы настроили параметры модели сегментации для оптимизации результатов подтвержденных пользователями измерений.

В *Таблице 1* показана точность, полученная для четырех подклассов программ.

Таблица 1 – Показатели точности автоматической сегментации

Класс	Tg1	Tg2	Tg3	TgR
Точность	0.81	0.69	0.80	0.73

Роль компонентов с открытым источником

Система, описанная в этой статье, нежизнеспособна без наличия инструментов с открытым источником. Во-первых, потому, что почти все компьютеры в данной архитектуре работают на операционной системе Linux, которая также была платформой развития и тестирования. Из компонентов, написанных на языке программирования C, скомпилированных с GNU Compiler Collection (GCC), и компонентов, написанных на более высокоуровневых языках, таких как python, perl или ruby, или простых скриптов в любом варианте Unix Shell ... были подготовлены различные ингредиенты, интегрированные в среду с открытым источником.

Для манипуляции аудио и видео материалом были успешно приняты *MJPEG Tools* [12]; управление рабочими процессами опирается на пару *Openflow* через *Zope* [13][14]; служба публикации построена как веб-приложение, использующее *Postgresql* [15] вместе с веб-серверами *Apache* http и *Tomcat* и поисковым механизмом *Lucene* [16].

Инструмент, использованный для сегментации говорящих в процессе редакторской сегментации – *mClust* [9], в для тематических категорий был принят *Categorizer* [17].

За концепцией инструментарий с открытым источником стоит среда с открытыми источниками, включая опыт участников, позволивший получить такую сложную систему с постоянными регулировками и требованиями к новым характеристикам.

В заключение

В статье мы дали обзор системы автоматической аннотации новостных программ ANTS, разработанной в RAI Research Centre в Турине. Сильная сторона принципа ANTS состоит в предложении пользователям хорошей глобальной производительности путем интеграции нескольких аналитических инструментов в один полностью автоматизированный продукт.

Документация элементов теленовостей должна давать практичные результаты с гораздо меньшей задержкой, чем для других телепрограмм. С ручным процессом аннотации – хоть и с помощью автоматического сбора и обнаружения планов – один элемент занимал пару рабочих дней, прежде чем стать доступным пользователям, а после внедрения ANTS информационная передача становится доступна для поиска на уровне отдельных сюжетов в течение двух часов после публикации.

Система ANTS сейчас используется в RAI для индексирования главных новостных редакций трех национальных каналов RAI1, RAI2 и RAI3. Пользователи могут делать прямые запросы в систему или подписаться на персонализированные RSS-каналы по определенным темам и получать уведомления при появлении нового контента. Служба будет прогрессивно расширяться для охвата и региональных новостей.

Другое интересное применение ANTS – для некоторых региональных итальянских администраций – мониторинг трансляций местных станций. В этом случае записываются и индексируются эмиссии всех соответствующих каналов. Затем работники администраций могут пролистывать контента для статистического анализа или верификации соответствия нормативам телетрансляции.

Будущая работа будет направлена на расширение редакторской сегментации других видов программ и развитие и интеграцию новых методов извлечения метаданных.

Официальное уведомление

Авторы хотят поблагодарить Laurent Boch и Daniele Airola из RAI CRIT за выдающийся вклад в дизайн и развитие системы ANTS.

Giorgio Dimino получил диплом по электротехнике Политехнического университета в Турине в 1987 г. В 1988 поступил в Исследовательский центр RAI – Radiotelevisione Italiana в Турине, работая в области обработки и архивирования цифрового звука и видео.

В его интересы входит проектирование автоматизированных цифровых архивов и применение информационной технологии в телепроизводстве.

Г-н Dimino руководит рабочей зоной метаданных, доступа и передачи 6-го проекта IST PrestoSpace. Также является активным членом EBU PMC и FIAT/IFTA..

Alberto Messina работает в Исследовательском центре RAI – Radiotelevisione Italiana в Турине: *Centro Ricerche e Innovazione Tecnologica* (CRIT). Участвует в нескольких внутренних и международных исследовательских проектах в области цифрового архивирования с особым акцентом на автоматизированное документирование и производство. Его сегодняшние интересы – от файловых форматов и стандартов метаданных до анализа контента и алгоритмов извлечения информации, что является сейчас его основным ориентиром. Также является автором различных технических и научных публикаций в данной области. Активно сотрудничает с Университетом Турина – департаментом вычислительной техники, что включает общие исследовательские проекты и преподавание.

Г-н Messina – активный член нескольких проектов EBU, в т.ч. P/TVFILE, P/ MAG и P/CP, и председатель проекта P/SCAIE, занимающегося техникой автоматического извлечения метаданных.

Roberto Borgotallo окончил факультет техники связи *Politecnico di Torino* в 1999 г. С 2001 г. работает в RAI – Radiotelevisione Italiana в научно-исследовательском департаменте (*Centro Ricerche e Innovazione Tecnologica*) в Турине. С самого начала участвовал в нескольких проектах по мультимедийному каталогу RAI под названием CMM, в области загрузки и трансформации метаданных.

В последнее время г-н Borgotallo работает в группе, разрабатывающей платформу автоматического извлечения метаданных, которая активно используется в RAI в экспериментальных целях и даже в производственной среде. Его основные профессиональные интересы – трансформация метаданных и сущности, системная интеграция и управление рабочими процессами.

Ссылки

[1] Trec video retrieval evaluation. Internet site: **http://www-nlpir.nist.gov/projects/t01v/**

[2] T. Chua, S. Chang, L. Chaisorn and W. Hsu: **Story boundary detection in large broadcast news video archives techniques, experience and trends**
In *Proc. of ACM Multimedia 2004*.

[3] W. Kraaj, A. Smeaton and P. Over: **Trecvid 2004: An overview**
In *Proc. of TRECVID Workshop 2004*.

[4] D. Eichmann and D.-J. Park: **Boundary and feature extraction at the university of Iowa**
In *Proc. of TRECVID Workshop 2004*.

[5] M.J. Pickering, L. Wong, and S.M. Rueger: **Anses: Summarization of news video**
In *Proc. of International Conference on Image and Video Retrieval (CIVR)*, 2003.

[6] T. Volkmer, S.M.M. Tahahoghi and H.E. Williams: **Rmit university at trecvid 2004**
In *Proc. of TRECVID Workshop 2004*.

[7] Y. Zhai, X. Chao, Y. Zhang, O. Javed, A. Yilmaz, F. Rafi et al: **University of central Florida at trecvid 2004**
In *Proc. of TRECVID Workshop 2004*.

[8] G.M. Qu'enot, D. Mararu, S. Ayache, M. Charhad and L. Besacier: **Clips-lis-lsr-labri experiments at trecvid 2004**
In *Proc. of TRECVID Workshop 2004*.

[9] Сайт: **http://www-lium.univ-lemans.fr/tools**

[10] M. De Santo, G. Percannella, C. Sansone and M. Vento: **Unsupervised news video segmentation by combined audio-video analysis**
In *MRCS*, pages 273–281, 2006.

[11] F. Brugnara, M. Cettolo, M. Federico and D. Giuliani: **A system for the segmentation and transcription of Italian radio news**
In *Proc. of RIAO, Content-Based Multimedia Information Access*, 2000.

[12] Сайт: **http://mjpeg.sourceforge.net**

[13] Сайт: **http://openflow.sourceforge.net**

[14] Сайт: **http://www.zope.org**

[15] Сайт: **http://www.postgresql.org**

[16] Сайт: **http://www.apache.org**

[17] Сайт: **http://search.cpan.org**