

11 Авторска справка за научните приноси на трудовете

През последните години работата ми е обхващала изучаването и предлагането на някои решения в различни части от цялостния процес по организация и обработка на информация, свързана с обекти от културното наследство. Постигнатите резултати са отразени в представените за конкурса научни публикации и могат да бъдат групирани в следните направления:

1. Автоматично извличане на метаданни от съдържанието на изображения, представящи художествени произведения.
2. Организация на цифрови хранилища за културно наследство.
3. Организация на достъпа.
4. Приложение на асоциативните класификатори при категоризация на обекти от културното наследство.
5. Откриване на връзки между зададено атрибутно пространство и комплексното описание на разглежданите обекти.

Представената книга "Access to Digital Cultural Heritage: Innovative Applications of Automated Metadata Generation" дава обща представа за съществуващата взаимосвързаност между различните области. Структурата на книгата и на отделните глави е предложена от мене и одобрена от останалите съредактори и съавтори.

1. Автоматично извличане на метаданни от съдържанието на изображения, представящи художествени произведения

Тези резултати основно са представени в дисертацията.

През годините в опитите за въвеждане на основа за преодоляване на различията между получаваната информация от човека и от компютрите се предлагат различни таксономии, описващи съдържанието на изображенията. В нашето проучване като основа се взема предложената от Hurtut (2010)¹ таксономия с известни промени в разпределението на характеристиките в групите, считайки че така по-точно се отчитат профилите на съответните характеристики. Самите групи са визуална, семантична и абстрактна. Визуалното пространство включва перцептивни примитиви (цвят, текстури, ъгли), геометрични примитиви (щрихи, контури, форми) и конструктивни примитиви (пространствено аранжиране, композиция). Семантичното пространство е свързано със значението на елементите, техния потенциал за семантично тълкуване. То се състои от семантични единици (обекти), пространствените връзки между тях (сцена, перспектива, дълбочина сигнали) и контекста (осветление, сянка). Абстрактните аспекти са специфични за изображенията на изкуството и отразяват културни влияния, специфични техники, както и емоционални експресии. Конкретните ми изследвания са били свързани с извличане на семантични и абстрактни характеристики на изображенията, базирани на цветовото съдържание на изображението. Тръгвайки от теорията на Itten (1961)² за цветовете контрасти, е направена

¹ Hurtut, T. 2D Artistic Images Analysis, a Content-based Survey. http://hal.archives-ouvertes.fr/hal-00459401_v1/

² Itten, J.: The Art of Color: The Subjective Experience and Objective Rationale of Color, Reinhold Publ. Corp. of New York, 1961.

класификация на цветовите хармонии и контрасти на основата на трите характеристики на цвета, които са най-близки до човешкото възприятие – нюанс, наситеност и светлота. Предложено е използването на нов цветови модел (комбиниращ качествата на три други модела), на чиято основа е изработено формално описание на хармониите и контрастите. Реализирани са функции за автоматично извличане на дефинираните хармонии и контрасти от цифровите изображения.

Реализиран е един подход за извличане визуални характеристики за класифициране живопис на базата на MPEG-7 дескриптори. MPEG-7 дескрипторите дават добро представяне на различни видове визуални характеристики, но са сложни структури и изискват специфични метрики за разстояние. Това затруднява прякото им използване в стандартните алгоритми за класификация. Нашият подход се състои от следните стъпки: (1) изображенията се разделят на незастъпващи правоъгълници с цел да се улови по-детайлна информация, (2) съответните парчета от изображенията се подлагат на клъстеризация за всеки MPEG-7 дескриптор; (3) с цел намаляване на размерността се използва векторно квантуване за дефиниране на уникална стойност на всяко парче, която съответства на номера на клъстера, към който принадлежи даденото парче. Проведени са анализи на значимостта на получените характеристики спрямо използваните MPEG-7 дескриптори, както и спрямо позицията на съответните парчета.

2. Организация на цифрови хранилища за културно наследство

Тези резултати основно са представени в [5], [6], [7] и [8] и в глава 6 от дисертацията.

Направено е изследване на съществуващите източници на информация за културно-историческото наследство в областта на художествените произведения – начинът им на зараждане, функциите, които изпълняват, прилаганите методи за усилване на високо-семантичното търсене в цифровите хранилища чрез комбиниране на текстовия подход за индексирание и автоматичната индексация на базата на съдържанията в изображенията [5]. Прототип на специализирана система за изучаване на анализ на изображението и извличане на характеристики е представен в [6].

В рамките на договор До02-308, финансиран от ФНИ (2009-2012) бе направен подробен анализ на текущото състояние на стандарти и спецификации за метаданни в областта на културното наследство и на съществуващи стандарти за хранилища [7]. Беше направено изследване на възприемането на основните елементи на цифровите библиотеки спрямо референтния модел за цифрови библиотеки DELOS DLRM – съдържание, потребители, функционалност, архитектура, качество и политики, обръщайки внимание на недооценяването на важните, но по-абстрактно възприемани елементи потребители, качество и политики [8].

Във връзка с разработването на методите и средствата за извличане, базирано на съдържанието на изображенията, е предложена архитектура и е разработена експериментална система APICAS ("Art Painting Image Colour Aesthetics and Semantics"). Като основа се използва моделът OAIS³ и се добавят функции за автоматично извличане, базирано на съдържанието, в рамките въвеждането и достъпа до цифровата библиотека. Предложените и реализирани функции за извличане на различните видове хармонии и контрасти целят намаляването на семантичната разлика между автоматично извличане на визуалните характеристики от ниско ниво и човешкото изразяване от високо ниво. Архитектурата и функционалните възможности на системата са представени в **глава 6 от дисертацията**.

³ Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems. Blue book - January 2002; Pink book - August 2009.

3. Организация на достъпа

Тези резултати са представени в [1], [10] и глава 5 от книгата.

Ефективното индексирание е от решаващо значение за изграждането и функционирането на големите бази от данни и търсачки. Направено е едно систематично проучване на различни методи за достъп като разширение на по-ранни анализи в тази област. Методите за достъп се класифицират в няколко категории: едномерни; многомерни пространствени; високо-размерни; времеви; пространствено-времеви методи за достъп [1]. В системите за извличане, базирано на съдържанието, индексиранието, наред с останалите два компонента – дизайна на характеристики и извличането, е структуроопределящ компонент.

През годините колективът, в който работя, използваше т.нар. многомерни номерирани информационни пространства, залегнали в дисертацията на колежата Красимир Марков. Основната характеристика на метода е замяната на имената с номера и използването на математически функции и адресни вектори за достъп до информацията. Най-общо предимствата на номерираните информационни пространства са възможност за изграждане на растящи пространствени йерархии от информационни елементи, както и мощност при изграждането на връзки между информационните елементи чрез използването на адресни индекси.

Беше създаден един алгоритъм за извличане на асоциативни правила, който използва възможностите на многомерните номерирани информационни пространства, наречен ArmSquare. Основен акцент при реализацията на ArmSquare е замяната на стойностите на атрибутите на инстанциите с цели числа, които са поредните номера на стойностите на елемента в съответните подредени множества от допустими стойности. Така всеки обект или комбинация от атрибути се описва чрез вектор с целочислени стойности, който може да бъде използван като координатен адрес в многомерно информационно пространство [10].

Бяха направени експерименти с множество от данни, включващи цветови атрибути, извлечени от експерименталната картинна база, разглеждана в дисертацията. В случая се използва транзакционна база, в която всеки ред съдържа името на художника, следвано от атрибутите, представени като двойки "<име на атрибут>=<стойност>". Използвайки възможността за групиране на данните (binning) по класовете са извлечени типични комбинации от характеристики за изследваните художници. ArmSquare може да се използва в първата фаза "генериране на правила" в алгоритъма на асоциативен класификатор. Основните изводи и възможностите за използване на ArmSquare в процеса на цифровизация и осигуряване на достъп до културното наследство са представени в **глава 5 от книгата**.

4. Приложение на асоциативните класификатори при категоризация на обекти от културното наследство

(Експериментална система PaGaNe и асоциативен класификатор PGN)

Резултатите са представени в [2], [3] и [9].

В рамките на колектива, в който работя, създадохме един класификатор, наречен PGN, който строи асоциативни правила, стремяйки се към запазване на максимална точност на произведените правила и използва мярка, измерваща степента на припокриване на въпроса със съответните правила (coverage measure) в етапа на разпознаване. Създадена е експериментална система, наречена PaGaNe, в която е реализиран предложеният класификационен алгоритъм. Проведени са

тестове върху експериментални бази от UCI repository⁴. Сравненията с различни класификатори, реализирани във WEKA⁵ показват ефективността на избрания подход [2].

Асоциативните класификатори по принцип обработват номинални атрибути. С цел осигуряване на възможността за обработка на множества, съдържащи и числови атрибути, като препроцесорна стъпка се реализират различни дискретизатори. След направен обзор в областта на съществуващите дискретизатори избрахме представители на различни типове, които реализирахме в експерименталната система PaGaNe –Equal-width, Equal-frequency, Fayyad-Irani и Chi-square. Проведените експерименти показаха ефективността на Chi-square спрямо останалите алгоритми. Сравнението на PGN, обучен с Chi-square с други класификатори, реализирани във WEKA, показва добри резултати в полза на PGN класификатора [3].

Идеологията и основната реализация на PGN класификатора принадлежи на колегата ми Илия Митов. Моята основна работа е в проектирането и функционалното структуриране на архитектурата на системата PaGaNe с цел: постигане на оперативна съвместимост на данните най-вече със системите WEKA, ORANGE и APICAS; отвореност за добавяне на нови алгоритми; модулност която да позволи комбинирането на различни техники като части от по-общ процес. За разлика от други области, където стандартите за описание на данни са навлезли в широка употреба, повечето от системите за извличане на знания все още използват собствени нотации, което затруднява силно интегрирането на процеси от различни приложения в обща задача.

Интересът ми към асоциативните класификатори е продиктуван от желанието за прилагането им при категоризация на обекти от културното наследство. Класификационните задачи в областта на културното наследство обикновено пораждат набори от данни с много класове. Асоциативните класификатори са удобни инструменти за извличане на знания в областта на културното наследство, използвайки техните възможности да вземат предвид специфични комбинации на стойностите на атрибутите.

Извършихме експерименти с множества от данни с много класове на примера на картинни колекции от различни периоди от Западно и Източно европейското изкуство. За целта използвахме атрибутно пространство, съдържащо автоматично извлечени характеристики от съдържанието на изображенията по методите, описани по-горе. PGN като представител на асоциативните класификатори беше сравняван с представители на други типове класификационни алгоритми, които строят подобен модел на разпознаване от гледна точка на възможността да се извлекат правила за дефиниране на класовете. За целта са използвани J48 (WEKA-реализацията на C4.5) като представител на Decision Trees и JRip (WEKA-реализацията на RIPPER) като представител на Decision Rules. Сравненията показаха конкурентоспособността на PGN при разпознаването на обекти от множества от данни с много класове [9].

5. Откриване на връзки между зададено атрибутно пространство и комплексното описание на разглежданите обекти

(Алгоритъм Meta-PGN)

Резултатите са представени в [4] и [12].

Разглежданите въпроси са на границата между процесите на клъстеризация, които разкриват структурата на атрибутното пространство и процесите на категоризация, които изграждат мостове

⁴ <http://archive.ics.uci.edu/ml/>

⁵ <http://www.cs.waikato.ac.nz/~ml/weka/>

между атрибутното пространство и понятията като елементи на по-сложно структурирано понятийно пространство.

Беше конструиран meta-PGN – разширение на стандартния класификационен алгоритъм PGN, чиято цел е комплексно обработване на многокатегориално зададено множество от данни и търсене на взаимовръзките между метаданните чрез откриване на зависимости в описателната част. Беше сметено, че представител на асоциативните класификатори е най-подходящ като основа за създаването на връзки с комплексното понятийно пространство поради способността им да откриват наличието на типични специфични комбинации от стойностите на атрибутите. Защото анализът на разпределението на цветовете в палитрите на различни художници показва, че някои артисти, например Гоген и Джото, използват подобни цветове в целия си живот и други, като Веласкес, имат голяма дисперсия в използваните палитри. В допълнение, това може да е сигнал, че работата на такива художници трябва да бъдат разделени на по-кратки периоди. Типичен пример за това направление е разнообразие от стилове Пикасо.

Анализът на съответствията между връзките, установени в атрибутното пространство и съществуващите връзки между понятията може да се използва като тест за създаване на адекватен модел на наблюдавания свят.

Дата: 15.01.2013

Подпис:

Резюмета на представените материали

Дисертация:

Ivanova, K.: A Novel Method for Content-Based Image Retrieval in Art Image Collections Utilizing Colour Semantics. PhD Thesis, University of Hasselt, Belgium, 2011.

Пред процеса на извличане на изображения стои голямо предизвикателство: той трябва да преодолее очевидната разлика между цифровите технологии, ограничени в рамките на улавянето на пикселите и човешките очаквания за възприемане на различните семантични, естетически и културни послания, които произведението изпраща на зрителя.

Цветът играе важна роля в анализа на картините. Възприятието му зависи от редица аспекти, започвайки от физическата природа на светлината, физиологичните особености на човешкото зрение, психологическите особености и социокултурните влияния, в които е било създадено произведението, както и в които се намира възприемащия посланието.

Успешните цветови комбинации се основават главно на закона, че човешкото око се чувства в равновесие само когато са налични допълващите се цветове.

Основната цел на тази работа е да се направи подробен анализ на съществуващите взаимни връзки в успешните цветови комбинации, както и да се формализират с цел прилагане на автоматичното извличане от цифровизираните произведения.

Предложихме използването на цветови модел (комбиниращ качествата на три други модела) спрямо трите основни характеристики на цвета – нюанс, наситеност и осветеност. На негова основа изработихме формално описание на хармониите и контрастите.

Предложихме примерна архитектура на система за извличане, базирана на съдържанието на изображенията като среда за прилагане и тестване на предложените алгоритми. Разработена е експериментална система APICAS.

Предложените и реализирани функции за извличане на различните видове хармонии и контрасти целят намаляването на семантичната разлика между автоматично извличане на визуалните характеристики от ниско ниво и човешкото изразяване от високо ниво.

Книга:

Ivanova, K., Dobрева, M., Stanchev, P., Totkov, G. (editors): Access to Digital Cultural Heritage: Innovative Applications of Automated Metadata Generation. Plovdiv University Publishing House "Paisii Hilendarski", 2012.

Основната цел на тази книга е да представи общ преглед на настоящите тенденции в областта на цифровизацията на културното наследство, като основният акцент е върху организирането на цифрово съдържание, генерирането на метаданни, както и на методи за обогатяване на откриването на ресурси. Книгата е резултат на успешното сътрудничество между екипите на различни български и чуждестранни институции, обединявайки резултатите от национални и международни изследователски проекти, насочени към разработване на специализирани технологии и инструменти, които биха били от помощ в усилията за цифровизация. Книгата се състои от пет глави. Тук ще дадем кратко обяснение на всяка от тях.

Глава 1 "Цифровизация на културното наследство - стандарти, институции, инициативи" прави въведение в областта на цифровизацията. Разглеждат се основните стълбове на процеса на създаване, съхраняване и достъп на културното наследство в цифровото пространство.

Подчертано е значението на метаданните в процеса на достъп до информация. Обсъждат се различни стандарти и схеми за метаданни, използвани в културното наследство. За осигуряването на достъп до цифровите обекти във виртуалното пространство, те се организират в цифрови библиотеки. Съвременните цифрови библиотеки се опитват да доставят по-богата и по-добра функционалност, ориентирана към потребителите. Главата се фокусира и върху някои инициативи на световно и европейско равнище, които значително променят процеса на цифровизация и организация на цифровите обекти в областта на културното наследство. През последните години, основният акцент в създаването на цифрови ресурси се измества от "фокусирани върху системата" към "фокусирани върху потребителите", доколкото повечето от въпросите са свързани достъпността и използваемостта от реалните потребители. Поради това използването на потребителските проучвания и участието на потребителите още в ранните етапи на проектиране и планиране на функционалността на продукта добива все по-голяма популярност.

Глава 2 "REGATTA – Регионален агрегатор на разнородни културни артефакти" описва прототип на агрегатор, представящ различни видове колекции – музейни сбирки, археологически обекти и недвижимо наследство. Избраният подход поддържа идеята за запазване на ценните национални паметници в европейското културно пространство чрез съхраняване на тяхната идентичност и специфика. Той е проектиран по стандартите на Europeana и характеристиките, посочени в българския регламент за създаване и управление на музейните фондове. Обсъждат се функционалната схема, моделът на данните и технологичните аспекти в REGATTA. Като експериментални данни в агрегатора са представени артефакти от фонд "Занаяти" от Етнографския музей в Пловдив, като е осигурена и възможност за провеждане на виртуални разходки в музея.

Глава 3 "Автоматизирано извличане на метаданни от арт изображения" се фокусира върху извличането от изображения, базирано на съдържание. През последното десетилетие е направен значителен напредък в посока на индексирание и анотиране на картини, представящи широк кръг от понятия. Областта на извличането от изображения трябва да преодолее голямо предизвикателство: тя трябва да се преодолее очевидната разлика между системата на човешкото възприятие, която се е развила генетично през хилядолетията, и цифровите технологии, които са ограничени в рамките улавяне и анализ на пиксела. Предизвикателството е още по-голямо когато се фокусира върху анализът на естетическото и семантичното съдържание на художествени образи. Извличането от изображения, базирано на съдържание, е технология, която помага за организиране на цифрови изображения, въз основа на тяхното съдържание. По този начин могат да бъдат извлечени разнообразни характеристики на различно ниво на концептуализация. Обикновено този процес се състои от три компонента: проектиране на характеристики, индексирание и извличане. Компонентата за проектиране извлича визуалните характеристики от изображенията. Компонентата за индексирание организира визуалната информация с цел ускоряване на обработката на заявките. Компонентата по извличане обработва заявката на потребителя и осигурява потребителския интерфейс. По време на този процес, централният въпрос е да се определят адекватното представяне на характеристиките и метриците за близост.

Глава 4 "APICAS – извличане, базирано на съдържанието на изображенията, в картинни колекции, използващи семантиката на цвета" предлага кратък преглед на теорията на цветовете от различни гледни точки. Обосновката за това е силната връзка на всяко произведение на изобразителното изкуство със сложната област на цветовото възприятие. Като отправна точка се разглеждат физиологичните основи на това явление. Направен е кратък исторически преглед на опитите за определяне на връзките между цветовете и взаимните им влияния. Визуалните характеристики от ниско ниво, представящи разпределенията на

цветовите характеристики са избрани като основа за изграждане на концепти от по-високо ниво. Направена е класификация на хармониите и контрастите в съответствие на теорията на Itten от гледна точка на трите основни характеристики на цвета - нюанс, наситеност и осветеност. Създадено е формално описание на определените хармонии и контрасти. Описан е и метод за извличане на локални характеристики, които улавят по-фина информация за използваните цветове и текстура, основан на разделянето на изображението на части и прилагане на векторно квантуване на MPEG-7 дескрипторите, изчислени за тези части. Програмната система APICAS е разработена с цел да осигури подходяща среда за реализиране на предложените алгоритми, както и за провеждане на експерименти. Разнообразието от приведените резултати от експерименти за използването на системата за различни задачи показва жизнеността на предложените функции като стъпка в прехода от Web 2.0 към Web 3.0.

Глава 5 "Автоматично генериране на метаданни и цифрово културно наследство" анализира възможностите за използване на някои автоматични методи за генериране на метаданни в областта на културното наследство. Методите засягат различни части от областта на извличане на знания, показвайки разнообразието от възможности за автоматизация на различните етапи. В началото е разгледан един подход за непряко извличане на пространствени данни чрез обучение на автомати с ограничен брой крайни състояния. Реализираната система InDES е тествана над извличане на пространствени метаданни от уебсайтове и показва обещаващи резултати. Това даде увереност, че такъв подход може да се използва за извличане на метаданни от описания на обекти и по този начин да бъдат използвани в процеса на миграцията от съществуващи по-ранни описания на обекти в случаите, когато тези описания са в неструктурирана форма. В следващата част е представен един метод за извличане на асоциативни правила, използващ възможностите на многомерните номерирани информационни пространства, наречен ArmSquare. Той е реализиран в средата за извличане на данни PaGaNe. Възможностите за извличане на често срещани комбинации могат да се използват за установяването на връзки между метаданни в рамките на дадена онтология. Асоциативните алгоритми се основават на подобни техники, но в областта на категоризацията. Асоциативният класификатор PGN също е част от системата PaGaNe. Той беше приложен в областта на анализ на семантични характеристики, извлечени от арт изображения на базата на съдържанието им. PGN-класификаторът успешно може да се използва за извличане на ресурси в рамките на управлението на данни и осигуряването на достъп до обектите в агрегаторите на културни артефакти.

Статии:

[1] Markov Kr., Kr. Ivanova, I. Mitov, St. Karastanev. Advance of the Access Methods. International Journal "Information Technologies and Knowledge", Vol. 2/2008, No.: 2. pp.123-135, ISSN: 1313-0455.

Ефективното индексване е от решаващо значение за изграждането и функционирането на големите бази от данни и търсачки. В статията се представя едно систематично проучване на различни методи за достъп като разширение на по-ранни анализи в тази област. Методите за достъп се класифицират в няколко категории: едномерни; многомерни пространствени; високо-размерни; времеви; пространствено-времеви методи за достъп. Статията включва описание на връзките между различните методи за достъп, както и позоваване на източниците, където са описани тези методи.

[2] Mitov I., Kr. Ivanova, Kr. Markov, V. Velychko, K. Vanhoof, P. Stanchev. "PaGaNe" – A Classification Machine Learning System Based on the Multidimensional Numbered Information Spaces. In Int.Conf. "Intelligent Systems and Knowledge Engineering" (ISKE2009), 27-28.11.2009, Hasselt, Belgium; Printed in World Scientific Proceedings Series on Computer Engineering and Information Science, No:2, pp. 279-286, ISBN: 978-981-4295-05-5.

В нашата работа сме създали един класификатор, PGN, който строи асоциативни правила, стремейки се към запазване на максимална точност (accuracy) на произведените правила и използва мярка, измерваща степента на припокриване на въпроса със съответните правила (coverage measure) в етапа на разпознаване. Създадена е експериментална система, наречена PaGaNe, в която е реализиран предложеният класификационен алгоритъм. Проведени са тестове върху експериментални бази от UCI repository. Сравненията с различни класификатори, реализирани във WEKA показват ефективността на избрания подход.

[3] Mitov I., Kr. Ivanova, Kr. Markov, V. Velychko, P. Stanchev, K. Vanhoof. Comparison of Discretization Methods for Preprocessing Data for Pyramidal Growing Network Classification Method. Presented at Int.Conf. i.Tech-2009, Madrid, Spain, 02-05.09.2009. Published in Int. Book Series "Information Science & Computing" – Book No: 14. New Trends in Intelligent Technologies, Sofia, 2009, pp. 31-39, ISSN: 1313-0455.

Асоциативните класификатори по принцип обработват номинални атрибути. С цел осигуряване на възможността за обработка на множества, съдържащи и числови атрибути, обикновено като препроцесорна стъпка се реализират различни дискретизатори. След направен обзор в областта на съществуващите дискретизатори избрахме представители на различни типове, които реализирахме в експерименталната система PaGaNe. Бяха реализирани Equal-width, Equal-Frequency, Fayyad-Irani и Chi-square дискретизатори. Проведените експерименти показаха ефективността на Chi-square спрямо останалите алгоритми. Установи се, че като цяло PGN, обучен с данни, обработени предварително с Chi-square постига по-добра разпознавателна способност спрямо останалите. Сравнението на PGN, обучен с Chi-square с други класификатори, реализирани във WEKA, показва много добри резултати в полза на класификатора PGN.

[4] Ivanova Kr., I. Mitov, Kr. Markov, P. Stanchev, K. Vanhoof, L. Aslanyan, H. Sahakyan. Metric Categorization Relations Based on Support System Analysis. In Proceedings of the VIIth Int. Conf. "Computer Science and Information Technologies", 28.09-02.10.2009, Yerevan, Armenia, pp.85-88, ISBN: 978-5-8080-0797-0.

В статията се разглеждат теоретични и практически аспекти на системите за класификация и техниките на обучение. Анализът на предметната област и анализът на класификационните схеми повдигат редица нестандартни въпроси, като например отношенията между категоризацията и метаданните и логико-комбинаторното структуриране и групирането на описателната част в множествата от данни. В статията се разглеждат няколко такива въпроси. Като илюстрация на примерен начин за разрешаване на разглежданите въпроси е направено разширение на стандартния класификационен алгоритъм в системата PaGaNe с цел комплексно обработване на многокатегориално зададено множество от данни и търсене на взаимовръзките между метаданните чрез откриване на зависимости в описателната част.

[5] Ivanova, K., Dobрева, M., Stanchev, P.: *Discovery and use of art images on the web: an overview. Third International Euro-Mediterranean Conference EuroMed 2010, 8-13.11.2010, Lemesos, Cyprus, Archaeolingua, pp. 205-211, ISBN 978-963-9911-16-1.*

Статията прави преглед на областта на откриване на ресурси в картинни колекции. Цифровизираните изображения на изкуството са важна част от културното наследство, представено в световната мрежа. Цифровизацията и достъпността в последното десетилетие основно се фокусират върху материалите от библиотеки, музеи, архиви и аудио-визуални архиви. Цифровизираните картини, както и цифрово-родените изображения са голямо предизвикателство за откриването и използването на ресурсите. Практиката показва, че подобри резултати в откриването на ресурси могат да бъдат постигнати чрез методи, които комбинират заявки върху метаданните с методи за извличане базирани на съдържанието. Статията представя еволюцията на хранилищата за цифрово картинно изкуство, след което представя накратко някои от основните методи за интелигентно откритие на ресурси в тази област, както и някои посоки за по-нататъшно развитие, което ще обогати още повече потребителския опит.

[6] Stanchev P., Ph. Ein-Dor, K. Ivanova, I. Mitov. *High Level Semantic Retrieval in Web Based Art Painting Digital Repositories. Proc. of Third Int. Sci. Conf. ISK 2010, Varna, 27-30.06.2010, pp.159-169, ISSN: 1313-4345.*

Статията представя, сравнява и предлага подобряване на съществуващите подходи към проблема за семантично търсене на високо ниво в цифрови картинни колекции. Представена е прототип на специализирана система за изучаване на анализ на изображението и извличане на характеристики.

[7] Станчев, П., Иванова, К., Койчев, И., Сомова, Е., Тотков, Г.: *Стандарти за мултимедийно съдържание. Глава 4 от Стандарти и спецификации на метаданни за е-документи, под ред. на Г. Тотков, Издателство на ПУ "П.Хилендарски", Пловдив, 2010, стр.116-156, ISBN: 978-954-423-650-2.*

Конкретно от автора са анализирани стандартите:

MPEG-7 – предназначен за интерфейсно описание на мултимедийно съдържание (Multimedia Content Description Interface), който осигурява стандартизирана технология, позволяваща описание и идентифициране на аудио-визуално съдържание в мултимедийни среди.

MPEG-21 – стандарт за споделяне на цифрови права, разрешения и ограничения за цифровото съдържание от създателя на съдържание към неговите потребители.

RDF – основен метод за концептуално описание или моделиране на информацията, съдържаща се в уеб-ресурси, които са с разнообразни формати и синтаксис.

Dublin Core – стандарт за описание на различни цифрови материали в уеб-пространството.

Разгледани са различните нива на оперативна съвместимост.

Разгледано е историческото развитие на уеб от 1.0 към 3.0.

[8] Сотирова, К., Иванова, К., Добрева, М.: Информационната грамотност в областта на цифровите библиотеки: проблеми и предизвикателства. Доклади на четвъртата нац. конф. Образованието в информационното общество, Пловдив, България, 26-27.05.2011, стр.25-34, ISSN: 1314-0752

Според референтния модел за цифрови библиотеки DELOS DLRM има шест основни области, които са застъпени в цифровите библиотеки – съдържание, потребители, функционалност, архитектура, качество и политики. Докладът разглежда представите за цифровите библиотеки на студенти-филолози в началото на изучаване на курс по тази дисциплина. Студентите свързват цифровите библиотеки със съдържание, функционалност и архитектура, без да си дават ясна сметка за останалите три области. Докладът анализира какви знания и умения трябва да се застъпят в обучението по тази дисциплина, като се има предвид, че технологичната му част остарява изключително бързо.

[9] Ivanova, K., Mitov, I., Stanchev, P., Dobрева, M., Vanhoof, K., Depaire, B.: Applying associative classifier PGN for digitised cultural heritage resource discovery. In Proc. of the 1st Int. Conf. "Digital Preservation and Presentation of Cultural Heritage", Veliko Tarnovo, Bulgaria, 11-14.09.2011, IMI-BAS, Sofia, 2011, pp.117-126, ISSN: 1314-4006.

Откриването на ресурси е една от ключовите услуги в цифровизираните колекции на културното наследство. Тя изисква интелигентно търсене в хетерогенно цифрово съдържание с големи обеми, което определя насоките на развитие на класификационните методи в последно време. Асоциативните класификатори са удобни инструменти за извличане на знания, използвани в областта на културното наследство, използвайки техните възможности да вземат предвид специфични комбинации на стойностите на атрибутите. Обикновено асоциативните класификатори дават приоритет на подкрепата пред доверието. Предложеният класификатор PGN подлага под въпрос този общ подход и се фокусира първо на доверието, като запазва само 100% доверителните правила. Класификационните задачи в областта на културното наследство обикновено пораждаят набори от данни с много класове. Това разнообразие е причинено от богатството на натрупаната култура през вековете. Сравненията на класификатора PGN с други класификатори, по-специално с JRip и J48, показват конкурентоспособността на PGN в разпознаването на масиви от данни с много класове на примера на картинни колекции от различни периоди от Западното и Източно европейското изкуство.

[10] Mitov, I., Ivanova, K., Depaire, B., Vanhoof, K.: ArmSquare: an association rule miner based on multidimensional numbered information spaces. 1st Int. Conf. on Advances in Information Mining and Management, IMMM-2011, Barcelona, Spain, 23-29.10.2011, ISBN: 978-1-61208-162-5, pp.143-148.

В статията се предлага един подход за извличане на асоциативни правила, който използва възможностите на многомерните номерирани информационни пространства, наречен ArmSquare. Основен акцент при реализацията на ArmSquare е използването на предимствата на този тип метод на достъп, т.е. възможността за изграждане на нарастващи пространствени йерархии на информационните елементи и задаването на междусистемни връзки между информационните елементи, съхранявани в информационната база. Това позволява промяна на търсенето с директен достъп до данните при добре структурирани задачи. Направените експерименти в различни области показват живостта на предложения подход.

[11] Hadjikolev, E., Vragov, G., Totkov, G., Ivanova, K.: REGATTA – Regional aggregator of heterogeneous cultural artefacts. Presented at the Sixth SEEDI Conference "Digitization of cultural and scientific heritage", 18-20.05.2011, Zagreb, Croatia. Published in Review of the National Center for Digitization, Belgrade, Serbia, Issue: 21, 2012, pp.8-18, ISSN: 1820-0109.

Статията представя проекта и реализацията на хранилище, имащо за цел да съхранява широк спектър от хетерогенни цифрови материали, представящи движимо и недвижимо културно наследство. Целта е да се даде възможност за популяризация на наличните ценности на местно ниво, спазвайки изискванията на Еуропеана с цел последваща агрегация. Подходът се основава на решения с отворен достъп. Проектът цели да запълни липсата на агрегатор за музейно съдържание и на недвижимо наследство в България поради специфичните условия за въвеждане на цифровизацията в българските институции за културно наследство, при която всяка институция сама за себе си търси решения на въпросите за цифровизация и осигуряване на достъп до цифровото съдържание.

[12] Ivanova, K., Mitov, I., Stanchev, P., Ein-Dor, Ph., Vanhoof, K.: Establishing correspondences between attribute spaces and complex concept spaces using Meta-PGN classifier. Proc. of the 2nd Int. Conf. "Digital Preservation and Presentation of Cultural Heritage", Veliko Tarnovo, Bulgaria, IMI-BAS, Sofia, 2012, ISSN: 1314-4006, pp.71-77

Статията представя един възможен подход за разширяване на задачата за класифициране чрез въвеждане на допълнителни метаданни, представящи различни страни на разглежданите обекти. Целта е анализът на съответствието между връзките, установени в атрибутното пространство и съществуващите връзки между понятията в концептуалното пространство да се използва като тест за създаване на адекватен модел на наблюдавания свят. Като възможен инструмент за създаване на тези връзки се използва разработеният от колектива алгоритъм Мета-PGN. Прилагането на този подход в областта на извличане, базирано на съдържанието, от картини предоставя инструмент за извличане на специфични комбинации от атрибути, представящи различни страни на художествени стилове, периоди и тематики и взаимовръзките между тях.

Дата: 15.01.2013

Подпис: