



GREENSTONE ЦИФРЛЫҚ КІТАПХАНАСЫ ҚАҒАЗДАН КОЛЛЕКЦИЯҒА

Иан Витен және Стефан Бодди

Компьютерлік ғылымдар факультеті
Вайкато Университеті, Жаңа Зеландия

Greenstone цифрлы форматтағы кітапханалық коллекцияны жасау және тарату үшін қолданылатын программа. Ол Интернет немесе компакт-дискілерде ақпараттарды ұйымдастыру мен жариялаудың жаңа тәсілін қолдануға мүмкіндік береді. Greenstone Вайкато университетінің жанында New Zealand Digital Library Project жобасы бойынша жасалынып, жетілдіріліп **ЮНЕСКО мен Human info** үкіметтік емес ұйымының қолдауымен таратылған. Бұл программа <http://greenstone.org> сайтында GNU General Public License шартымен орналасып жалпыға бірдей пайдалануға болалыды.

Біз өзіміздің программалық жабдық жұмысының сізге еш қолайсыздық тудырмайтындығына сенімді болғымыз келеді. Сізде Greenstone-де жұмыс істеу кезінде сұрақтар туындаса бізге хабарласыңыз greenstone@cs.waikato.ac.nz

Бұл нұсқауда қағаз құжаттардан CD-ROM коллекциялар жасау толық түсіндіріледі. Мұнда сканирлеу мен символдарды оптикалық танудың процедурасы мен экономикалық жағы түгел сипатталған, демек сіз Greenstone программалық жабдығын қолданып мәтінді дұрыс форматқа көшіре аласыз. Сондай-ақ коллекцияда жиналған материалды түзетіп, коллекция жасау үшін Органайзерді қалай қолдану керектігі сипатталған. Бұл тегін Greenstone-мен бірге таратылатын Windows-қа арналған программа.

Ілеспелі құжаттар

Greenstone арналған ақпараттардың толық жинағы 4 бөліктен тұрады

- Greenstone цифрлы кітапханасын орнату бойынша басшылық
- Greenstone цифрлы кітапханасын қолданушыға нұсқау
- Greenstone цифрлы кітапханасын жасаушыға басшылық
- Greenstone цифрлы кітапханасы: Қағаздан Коллекцияға дейін.

Алғыс

Greenstone программасы көптеген адамдардың бірлескен еңбегінің жемісі. Роджер МакНаб пен Стефан Бодди басты жасаушылар. Дэвид Бэйнбридж, Джордж Букенан, Хон Чен, Элке Дюнкер, Карл Гутвин, Геофф Холмс, Джон Макферсон, Крейг Невилмэннинг, Гордон Пэйнтер, Бернхард Фарингер, Тодд Рид, Билл Роджерс және Стюарт Йетис өз тараптарынан еңбек сіңірді. Бұл жобаның басқада мүшелері Марк Апперли, Сэлли Джо Каннингем, Стив Джонс, Те Така Кигэн, Мишел Лутс, Малика Махоуи және Ллойд Смит программа дизайнын жасауда шабыттандырушы болды. Біз сонымен қатар осы программаға MG, GDBM, WGET, WV, PDF2HTML, PERL қосымшаларын жасауға еңбек сіңіргендерге алғысымызды білдіреміз.



Мазмұны

Нұсқау жайлы	ii
1 Кіріспе	1
2 СКАНЕРЛЕР ЖӘНЕ СКАНИРЛЕУ	3
2.1 Сканерлер	3
Қымбат емес столға қоятын сканерлер	3
Қағазды автоматты түрде беретін сканерлер	4
Түрлі-түсті сканерлер	4
Кәсіби дуплексті сканерлер	4
Сканирлеуге арналған программалар	5
2.2 Құжаттарды дайындау	5
2.3 Сканирлеу процесі	5
Сапаны бақылау	6
Құжаттарды белгілеу үшін ұсынылатын ережелер	6
2.4 Өнімділік пен ресурстар	7
Сканирлеудің бағасы	7
3 СОР: символдарды оптикалық тану	11
3.1 СОР процесі	12
Сапаны бақылау	12
Кестелер	13
Бейнелер	13
Арнайы материал	14
3.2 Өнімділік пен қол жетерлік ресурстар	14
Интенсивті ОРС	15
Өнімділік	16

3.3 СОР-дың Алтернативасы	17
Қолмен қайта теру	17
Бейнелер файлы	17
3.4 Сканирлеуді СОР-мен бірлестіру	18
4 Үш мысал: 1000 нан 100.000 бетке дейін	19
4.1 Шағын коллекция: 500 беттен 1000 бетке дейін	19
4.2 Ұйымның барлық басылымы: 5000 бет	20
4.3 Шағын кітапхана: 100.000 бет	20
5 ЭЛЕКТРОНДЫ КОЛЛЕКЦИЯ ЖАСАУ	23
5.1 Коллекция жасаудың тәсілдері	23
5.2 Органайзер	24
Органайзерді орнату және қолдану	25
Құжат моделі	26
Органайзерге шолу	27
5.3 Құжаттар файлына тәгтер жасау	39



Greenstone - программалық жабдықтарын жасаудағы мақсаттардың бірі - ол әртүрлі институттарға, ұйымдарға, БҰҰ агенттіктеріне, үкіметтік емес және коммерциялық емес ұйымдарға, сонымен қатар үкіметтерге Интернетте орналастыруға және CD-ROMда сақтауға болатын ақпараттық коллекциялар жасау мүмкіндігін беру.

Төменде коллекция жасаудағы кәдімгі процедура сипатталған:

- i. Бастапқы құжаттарды тандау.
- ii. Авторлық құқықты иемденушілерден бұл құжаттарды қолдануға рұқсат алу.
- iii. Материалдық құжаттарды цифрлы форматқа түрлендіру үшін Сканирлеу мен СОТ.
- iv. Барлық құжаттарды Greenstone да импорттауға болатын бірдей форматқа түрлендіру (HTML және Microsoft Word форматтары дұрыс болар еді, программаға қосымшалар "plugin" көмегі арқылы басқада форматтарды қолдануға болады (қолданушыға арналған нұсқауды қара).
- v. Цифрлы құжаттардағы тарау, параграфтар және суреттерді маркерлеу.
- vi. Коллекцияларды оптималды құрылымдалған цифрлы кітапханаға ұйымдастыру.
- vii. Цифрлы кітапхананы жасау.
- viii. Коллекцияны CD-ROMға жазу және тарату және\немесе Интернетте жариялау.

Цифрлы коллекцияны жасау үшін, сәйкесінше барлық басылымдарды цифрлы форматқа түрлендіру керек. Егерде кітаптар мен құжаттардың қатты көшірмелері ғана бар болса, онда оларды сканирлеп, компьютер оқи алатындай түрге келтіру қажет (iii қадамы). Әдетте оны СОТ (Символдарды оптикалық тану) жүргізу арқылы жасайды, ал басқа жағдайларда жәй қайта теру арқылы. Бұл процесс 2-4 тарауларда

2 КІРІСПЕ

қарастырылады. V. қадам құжаттардың жекелеген бөліктерін таңдап алып, оларды кітапханада белгілі жерге орналастыру мүмкіндігін береді, vi. қадам құжаттарға атаулар көрсеткіші, түйін сөздер және библиографиялық мәліметтер сияқты белгілі атрибуттарды сұраным жасау мен кітапханадан іздеу жүргізу үшін береді. Бұл қадам Greenstone программалық жабдығына енетін Органайзер програмасын бүтіншегісіне жеткізіп сипаттайтын v. Тарауда қарастырылады.

Бұл нұсқау сондай-ақ қатты көшірмелерден цифрлы коллекция жасағандағы түзету процесіне қажетті көптеген мәселелерді талқылайды. Оқуды жалғастырудың алдында біз сіздерге келесі сұрақтарға жауап берулеріңізді сұраймыз:

- Коллекцияңызды жасаудағы мақсатыңыз не?
- Қандай топтағы адамдарға арнаған?
- Қаншалықты көлемді – локальды, аймақты немесе ғаламдық?
- Қанша құжат бар?
- Қанша бет бар?
- Қанша графикалық нысан бар?
- Коллекцияда шектеулі адамдар тобына арналған бөліктер бар, ал басқа бөліктері жалпы барлық қолданушыларға арналған ба?
- Бастапқы құжаттарға электронды формата қол жеткізуге бола ма?
- Егер болса, онда қандай формата? (PDF файлдардың толық мәтінді цифрлы форматтың эквиваленті еместігін және де тек беттердің көшірмесі екендігін атап өту керек).
- Бастапқы құжаттардың авторлық құқықтарының статусы қандай?
- Авторлық құқықтарды кім иемденген?
- Дәл осы топтағы адамдарға арналған өнімдері бар басқа ұйымдар бар ма?
- Сіздің басқа топтармен бірлесіп істегіңіз келе ме?
- Бұл жобаға бөлінген бюджет қандай?
- Үйлестіру, түзету, сканирлеу мен программалау үшін қол жететін адам ресурстары қандай?
- Қанша компьютерге қолыңыз жетеді?
- Қанша CD-ROM жасағыңыз келеді?
- Олар ақылы ма немесе ақысыз бола ма?



Сканерлер және Сканирлеу

Қағаз құжаттарды цифрлы коллекцияға түрлендірудің бірінші қадамы, ол құжаттардың барлық беттерін сканирлеу. Келесі қадам символдарды оптикалық тану процесін жүргізу (СОТ), ол үшін бастапқы құжаттардың жоғары сапалы және анық болғаны өте маңызды. Цифрлы форматқа түрлендіру процесі 300 dpi (дюмдегі нүктелер) анықтықта жұмыс істей алатын сканерді қажет етеді. Сканирлеудің көп бөлігін ақ-қара режимінде жүргізуге, ал түрлі түсті суреттер бар болса, онда оларды түрлі түсті сканермен сканирлеу қажет. Көпшілік жағдайда кітаптардың мұқабалары түрлі түсті әдемі безендірілген, оларды түрлі-түсті режимде сканирлеу керек.

2.1 Сканерлер

Сканерлер бағасы жағынан, көлемі мен формасы жағынан да әртүрлі. Олардың бағасы кәдімгі столдық сканерлер 100\$-дан ал Bell&Howll өндірушілері ұсынатын үлкен индустриалды сканерлер 50000\$ ға дейін барады. Әртүрлі сканерлерді ұсынатын көптеген Интернет сауда беттері бар. Сканер іздесеңіз Google, Altavista, Yahoo іздеу жүйелерін пайдаланыңыз. Сканирленген құжат әдетте TIFF немесе BMP (Bitmap image) форматында сақталады. TIFF IV сығылған формасы қолдануға ең жақсы формат болып табылады. Сканирленген орташа бет осы форматқа кешірілгеннен кейін 50 килобайт, ал сол бетті BMP форматында сақтасақ 2 Мб (мегабайт) орын алады.

Қымбат емес столға қоятын сканерлер

Қымбат емес столға қоятын сканерлер, ең арзан көпшілік сатып ала алатын сканерлер болып табылады. Мынадай көптеген сауда белгілері бар: HP, Agfa, Acer, т.т. Олардың бағалары 100 ден 300 доллар аралығында. Олардың көмегімен ақ-қара және түрлі-түсті суреттерді сканирлеуге болады.

¹Осы құжатта қолданылған ақша саммалары 2001 жылғы АҚШ долларында

4 СКАНЕРЛЕР ЖӘНЕ СКАНИРЛЕУ

Бағаларының төмендігі, әрбір қолданушының осындай сканерлерді иемденуіне мүмкіндік береді.

Олардың жағымсыз жақтары – ол сапасының орта деңгейде болуы, баяу сканерленуі, жылы жағдайларда сенімсіз болуы, және жиі істен шығуы. Беттер қолмен және әрбіреуі жеке-жеке сканирленуі. Әр бетті дұрыс орналастырып салу керек. Мұндай сканерлерде өнімділік өте төмен. Өндірушілердің 1 бетті 1 минутан аз уақытта сканирлеуге болады дегеніне қарамастан, практикада 1 сағатта 12 беттен аспайды. Сонымен бірге сканирлеу процесі компьютерлік ресурстарды толығымен алады.

Сондықтанда мұндай сканерлерді шағын беттік үлкен емес жұмыстарды ұдайы сканирлеуді жүргізсе айына 200-400 беттен аспайтын жұмыстарды жасауға ыңғайлы. Сканермен толық ставкада істейтіндер 1000-2000 бет жасайды.

Қағазды автоматты түрде беретін сканерлер

Қағазды автоматты түрде беретін сканерлердің бағасы 500-1200 доллар манайында. Сканерге 10-15 бетті бірден салып оларды бір уақытта сканирлеуге болады: сондықтан да, оператордың ұдайы аппарат жанында тұруының қажеті жоқ. Бұл өнімділікті күніне 150-200 бетке дейін арттыруға мүмкіндік береді. Бұл сканерлер төзімдірек және де ұзақ уақыт жөндеусіз (30000 -50000 бет сканирлегеннен кейін) істейді.

Олардың көпшілігі беттің бір жағының ғана сканирленуі, екінші жағын сканирлеу үшін бетті төңкеріп салу керек. Бұл қиыншылықтарды жиі тудырады, себебі автоматты түрде қағаз берушілер әркез қиындық туғызады және беттер жиі ішінде тұрып қалады.

Бұл сканерлер айына 1500-3000 бет өндеуге ыңғайлы.

Түрлі-түсті сканерлер

Сканирлеудің кез келген амалы түрлі-түсті суреттерге тап болуы мүмкін, сондықтан түрлі-түсті сканерді дайындықта ұстаған жөн. Әдетте кез-келген басылымның 5% дан азында мұқабасы мен қоса қандай да бір түрлі-түсті сурет болады. Сондықтанда арзан столға қоятын түрлі түсті сканер болғаны дұрыс. Олардың анықтық шешімі 600 dpi ден аз болмағаны дұрыс.

Кәсіби дуплексті сканерлер

Кәсіби сканерлер сенімді, мықты және күніне 2000-10000 бет аумағында сканирлеуге қабілетті.

СКАНЕРЛЕР ЖӘНЕ СКАНИРЛЕУ 5

Оларда қағазды автоматы түрде беру жүйесі бар, ол 50-200 беттік бумаларды өңдей алады. Ең жақсы және жылдам сканерлер – ол дуплексті сканерлер, олар бір уақытта құжаттың екі жағында сканирлейді.

Бұл сканирлерге өте күшті кемінде 10-20 Gb қатты дискісі бар компьютерлер қажет. Олардың бағасы 5000-50000\$ аралығында. Мысалы, Canon DR-6020 дуплексті сканері 5000 доллар тұрады, ол екі жақты құжаттар мен жұмыс істей алады.

Ол күніне 2000 бетке дейін сканирлеуге қабілетті және 600 000-800 000 бет сканирлегенге дейін жұмыс істейді. Bell&Howell and Fujitsu сканерлері 10000-50000 доллар аралығында, олар миллиондаған құжаттарды алғашқы жөнделімге дейін өңдей алады. Micro-fishe сканерлері 15000\$ жартылай автоматты түрі, 80000\$-ға дейін толық автоматты режимде істейтіні.

Сканерлеуге арналған программалар

Әр сканердің өз программалық жабдығы болады, сондықтанбұл программалық жабдықты сіз өз компьютеріңізге орнатуыңыз керексіз. Кейбір программалардың computer card-тары бар оны сканирлеу процессін жеделдету үшін орнатқан жөн.

2.2. Құжаттарды дайындау.

Құжаттарды сканирлеуге дейін дайындап қою қажет. Олардың шаңын қағып, су құжаттарды кептіріп, қыстырғыштарын алып беттерін түзету керек.

Әр кітаптың тігісін тиянақтылықпен алу керек. Сканирлеуге қолданған кітаптарыңыздың көпшілігін қайта тігу керек болады, сондықтан тігіс сөткенде ұқыпты болыңыз.

20-дан артық беті бар құжаттар үшін, бұл жұмысты керекті жабдыктары бар арнайы ұйымдарға бергеніңіз дұрыс.

2.3. Сканирлеу процесі

Сканермен бірге таратылатын программаны қолданып, сканирлеген соң сіз Bitmap немесе TIFF форматындағы цифрлы құжат аласыз.

6 СКАНЕРЛЕР ЖӘНЕ СКАНИРЛЕУ

құжаттарды қатты дискіде кез келген атпен сақтау қажет.

Сапаны бақылау

Сканирлеудің мақсаты – сканирленген құжаттарды символдарды оптикалық тану COT үрдісі арқылы (OCR, optical character recognition) WORD немесе HTML форматындағы құжаттар алу немесе құжаттар бейнесін PDF форматында алу. Кез келген жағдайда бастапқы құжат бейнесінің сапасы өте маңызды. Егер бейне сапасы стандарттан төмен болса, онда олардың бейнесі нашар және компьютердің жадысын көбірек алады. Бейненің сапасы символдарды оптикалық тану (COT) үрдісіне өте қатты әсер етеді. Нормадан төмен сапа жағдайында, өнімділігі 40%-ға төмендейді. COT, қатты көшірмелерді цифрлық форматқа түрлендірудің бүкіл үрдісінің құнының 90%-н құрайды, сондықтан сканирлеу сапасы соңғы шығындарға қатты әсер етуі мүмкін.

TIFF форматындағы файлдар сапасын, сканирлеу үрдісін қағаздың өртүрі үшін күйіне келтіру жолымен, программадағы күйіне келтірулерді өзгерту жолымен арттыруға болады. Айтарлықтай жұқа қағаз, басқасына қарағанда өзгеше күйіне келтіруді қажет етеді: Контрасты сканирленетін құжаттың басылу сапасына байланысты күйіне келтіру керек.

Алдымен бастапқы материалды қағаздары ұқсас және басылу сапасы ұқсастарын біріктіріп топтарға бөледі. COT-ң сынақтық үрдісін бірінші топтың қағаздарының бірнеше данасына жүргізіп оптимальды күйіне келтірулерді анықтаңыз. Содан кейін бұл топтағы материалдардың бәрін қорықпай-ақ сканирлеуіңізге болады.

Құжаттарды белгілеу үшін ұсынылатын ережелер

Әрбір кітап не құжатқа өз нөмірін немесе кодын беріңіз, ол осы құжаттағы барлық TIFF бейнелері бар папкінің аты болады. Операциялық жүйеге байланысты (DOS, Windows, Linux, Unix, т.т.) файл аттарында 8-ден 128 символға дейін қолдануға болады. Біз 8-16 символдан асырмауды ұсынамыз. Бірінші бес әріп құжат атын символдандырады да, одан кейінгі әріптер тіл кодын, ал қалған цифрлар – белгілі бір бетті білдіреді. Мысалы, u7548e.tif аты u7548 кодымен ағылшын тілінде жазылған 12-беттегі TIFF форматындағы суретін білдіре алады.

Сканирленген жұмыстарға арнап қатты дискіде бір директория жасап, оны

мысалы ScanJobs деп атаңыз.

Содан кейін әр тапсырма үшін ішкі директория жасаңыз. Ішкі директорияның ішінен әр басылым мен кітап үшін u7548 жоғарыдағы мысалда тағыда бір ішкі директория жасаңыз. Құжаттағы барлық TIFF бейнелерді, түрлі-түсті суреттерді қоса осы папкіде сақтаңыз.

2.4 Өнімділік пен ресурстар

Сіз сканирлеу процедурасының бүкіл масштабын әсіресе COT үрдісін дұрыс бағалауыңыз керек. Ең дұрысы сканирлеу мен COT екеуін толық тәуелсіз амалдар деп есептеу. Олардың әр қайсысы бойынша жекелей тиімді шешім қабылдауы қажет.

Мыналарды сканерлер мен компьютерлерді сатып алудың алдында ойланған жөн: қажетті бөлме мен жұмыс күші, жұмысшыларды оқыту; жалақы; сканирлеуге қажетті құжаттардың бастапқы және соңғы беттері; мерзімі; бұл құжаттардың үшінші жақтағы біреулерге берілу мүмкіндігі.

Сканирлеудің бағасы

Қабылдайтын шешімдердің, ең маңыздысы – ол сканирлеуші аппаратураны сатып алып, өзіңізге сканирлеу керек пе, жоқ әлде сканирлеуді маманданған компанияларға тапсыру керек пе. Міне ойланатын басты пунктер:

Сканирлеуді аяқтаудың соңғы мерзімі

Беттердің жалпы саны

Сканирлеушілердің жалақысы

Сканирлеумен айналысатын адамдар жұмыс сапасына жоғары талап қоятын, икемді және жауапкершілікпен қарайтын болуы керек.

Кәсіби компаниялардағы сканирленген бір беттің орташа бағасы 0.06\$. Бұған бетті дамушы елден дамыған елге 0.03\$ елдің өзінің ішінде әр бетке 0.015\$ тасымалдау қызметі үшін қосуы керек.

1-кестеде Әр түрлі сканерлерді қолданып сіз жасаған бұл амалдардың құны бағаланады. Бардық цифрлар шамалап берілген. Олар шамалап, авторлар тәжірибесіне негізделген. Алғашқы үш бағанда жұмыс күшінің бағасы берілген. Біріншісінде өнімділік бет/айына, толық ставкадағы жұмыс алынған. Бір бетке арналған Адам-сағаттарға қажет ресурстар бір айдағы жұмыс сағаттарын бет/айына өнімділігі арқылы табылып екінші бағанда берілген. Екінші бағанда бір айда 180 жұмыс сағаты бар деп алынған.

8 СКАНЕРЛЕР ЖӘНЕ СКАНИРЛЕУ

Кесте 1 Сканирлеудің бағасы

	Көлем (беттер / айына)	Сағ/бет (180 сағат айына)	Баға/бет (айына 4\$/сағ)	Сканер бағасы	Бірінші жөнделімге дейінгі жү- мыс істеу қабілеті (бет)	Кәсіби компания атқаратын көлем (бетіне 0.06\$ болғанда)
Беттік столға қоятын сканер	2500	0.072	\$0.288	\$300	7000	5000
Бетті автоматты түрде беретін сканер	8000	0.0225	\$0.09	\$800	30000	13000
Кәсіби: low-end duplex	40000	0.0045	\$0.018	\$6000	600000	100000
Кәсіби: high- end duplex	150000	0.0012	\$0.0048	\$50000	8000000	833000

Бір беттің бағасын анықтау үшін толық сағаттық жалақыны екінші бағанға көбейтіңіз. Мысалы, үшінші бағанда сканерге шығынды 4\$/сағ ескермей, еңбекті жалдау жолымен сканирлеудің бағасы берілген.

Бұл есептеулерде сканер жеткілікті көп көшірмелерге өзінің бастапқы құнын өтеу үшін қолданылалы деп есептеледі. Соңғы үш бағанда сканердің өзінің бағасы бойынша көп мәлімет береді. Біріншісі сканердің сатып алынған бағасын, келесі жұмыс істеу қабілетінің жуықша ұзақтығын. Соңғысы беттің бағасы 0.06\$ дан болғанда компанияны жалдап сканирлеуге болатын беттер санын көрсетеді.

Әрине сканерді таңдауға басқада факторлар әсер етеді: ақшаның бар болуы, басқалардан тәуелділікті минимизациялау қажеттілігі, кітапханалармен кітаптарды кітапханадан тыс шығармай-ақ сканирлеуге рұқсат беретін шарттар т.т.

Жоғарыда сипатталған кесте шығындарды өтеу үшін сканирлеуге қажетті беттер санына жуықша баға береді. Ұйымға 800000 бет сканирлеу керек болған жағдайлар өте сирек. Мұндай масштабта, құрал жабдықтарды ұстау мен оны жалға беру арқылы шығындарды өтеу сияқты күрделі мәселелер пайда болады, олар бұл бөлімде талқыланбайды.

Сканирлеу бизнесін әсіресе дамушы елдерде дамыту өте тартымды коммерциялық мүмкіндік болып көрінуі мүмкін. Есіңізде болсын,

СКАНЕРЛЕР ЖӘНЕ СКАНИРЛЕУ 9

құжаттар сканирленіп болғаннан кейін, сіздің клиенттеріңіз онымен сіздің қарым қатынасыңыз қаншалықты жақсы болғанына қарамастан ол сізге енді қайтіп ешқашанда сол құжаттарды сканирлеуге тапсырыс бермейді. Коммерциялық жағынан бұл бизнес интенсивті маркетингтік зерттеуді талап етеді. Біз үкіметтік емес ұйымдар мен коммерциялық емес ұйымдарға мұндай бизнеспен нарықты түбегейлі зерттемей және жақсы ойланып бизнес жоспар жасамай тұрып айналыспауды сұраймыз.

Қорытындыда, егер 1000 нан 50000 бетке дейін сканирлеу қажет болса, онда ең жақсысы оны сканирлеуші компанияға тапсырған дұрыс. Кәсіби low-end сканер, егер сіз 100000 бет сканирлесеніз өзінің құнын өтейді. Егер сіз осындай сканер алам деп шешсеніз, онда ең дұрысы басқа үкіметтік емес ұйыммен немесе кітапханамен бірлесіп алған жөн.

10 СКАНЕРЛЕР ЖӘНЕ СКАНИРЛЕУ



COT: Символдарды оптикалық тану

Символдарды оптикалық тану немесе COT сканирленген цифрлық бейнені мәтінге трансформациялайды. Бастапқы материал TIFF немесе Bitmap форматындағы цифрлы бейне – оның сапасының жақсы болғаны дұрыс . COT тан өткеннен кейін сіз RTF, Word, HTML форматындағы файлды өз қалауыңыз бойынша ала аласыз.

Міне, қағаздағы құжаттарды цифрлы форматқа көшіруде қолданатын қадамдар:

1. Сканирлеу
2. Талдау
3. Тану
4. Бейнелер мен кестелерді сканирлеу

Осы қадамдарды орындап сіз алынған файлдардың сапасын бақылап, оларды сәйкес папкілерде сақтай аласыз.

Нарық базарында бағасы 100-400 долларлық жақсы COT программалар жеткілікті.

Дистрибьютерлер тізімімен басқада барлық ақпарат өндірушілердің Интернет сайтында орналасқан. Автордың тәжірибесіне қарағанда олардың арасындағы қолдануда ең оңайы Fine-Reader мен Omnipage. Fine-Reader ең арзаны, оның бағасы бар болғаны 100\$. Ол икемді мүмкіндіктер мен әртүрлі тілдердің ең көп санын ұсынады.



12 СОТ: СИМВОЛДАРДЫ ОПТИКАЛЫҚ ТАҢУ

Сізге сканирлеу мен СОТ-ты өз күшіңізбен жүргізесізбе, жоқ әлде осы салада маманданған компанияға тапсырасыз ба, соны шешуіңіз қажет. Бұл процесті өз күшіңізбен жүргізу үшін сізге сканер, СОТ программалық жабдығы, СОТ пен жұмыс істеу машықтығын дамыту, жұмысты сапалы орындайтын жұмыс күші қажет.

3.1. СОТ процесі

Сот процесі, СОТ тың барлық программаларында өзгешеленеді, таңдалған программалардың кез-келгені жақсылап оқып үйренуді қажет етеді. Әрбір программаны пайдалануға арналған нұсқаулар бұл процесті жете түсіндіреді. СОТ процесінің төрт пункті ерекше зейін аударуға тұрарлық: сапаны қадағалау; кестелер, суреттер мен әдейі арналған формулалар сияқты материалдар, шет тілі және т.т.

Сапаны бақылау

Біз қайта-қайта сіздердің зейіңдеріңізді сапаны бағалауға аударғымыз келеді. Сапаны бақылауды құжат жазылған тіл, ана тілі болатын немесе осы тілді жетік білетін адамдарға тапсырған жөн. Ең жақсы жұмыскерлер, ол оқушылар мен студенттер. Себебі жастар мұндай жұмыста егде адамдарға қарағанда зейіндірек және жинақтылау бола алады.

Әдетте сапаны бақылаудың төрт кезеңі бар. Біріншісі СОТ процесі мен бір уақытта жүргізіледі. СОТ тың әрбір программасының ішінде орфографияға тексеру жүргізіліп жазылуы күмәнді сөздердің астын сызады. Екіншісі СОТ үрдісі аяқталғаннан кейін мәтінді жалпы тексеру. Бетті, абзацты, тарау аттарын және т.б. тастап кету сияқты қателер жиі кездеседі. Барлық беттердің барлығына жалпы шолу жасап және тақырыптарды, тарау аттарын, абзацтар мен кестелерді тексеру қажет.

Үшінші кезең Microsoft Word-та орфографияны тексеру. Word-та әдетте орфографиялық сөздіктер СОТ программаларына қарағанда толығырақ. құжатты Word-қа импорттап, орфографиялық тексеру жүргізу арқылы қателердің көбін түзетуге болады. Мәтіннің осындай түріне тән комплексті сөздер мен терминдерді олардағы қателерді тексеру үшін тексеру параметрлеріне енгізуді ұмытпаңыз.

Ең соңында, біткен құжат, толық кітап түзумен, орфографияны тексеруші, кесте мәселелерін, бейнелер мен аяқталған мәтіннің жалпы түрі мен айналысатын басқа адаммен тексерілуі қажет.

Тек осыдан кейін ғана электронды кітапты таратуға болады.

Кестелер

СОТ программалары кестелермен нашар істейді. Бұдан өрі, кестелерді тексеру қиын. Оларда көп цифрлар мен нүкте, үтір сияқты символдар бар, олар программаның қателесуімен басқа бағанға немесе жолға көшірілуі мүмкін. Сондықтан кестелер зейінді, көп жұмысты, шыдамдылықты және сапаны бақылауды талап етеді.

Біріншіден, кестелермен, бейнелер сияқты істеуге болады. Бұл кесте бейнелерін ақ-қара сканирлеу мен оларды құжаттың керек жеріне осы формада қоюдан тұрады. Бұл ең оңай жол.

Ешқандай қателер болмайды және барлық жұмсалған уақыт тек бейнелерді жасауға кетеді. Сондада, алынған кестелер бейнесі компьютерлік жадыны көп талап етеді. Сондай-ақ компьютер экранының шешуші қабілеті, үлкен кестелерді экранға шығарғанда жеткіліксіз болуы мүмкін. Егер сіздің, кестенің бәрін экранға орналастырғыңыз келсе, онда экранның шешуші қабілеті жеткіліксіз болуы мүмкін. Егер кесте өте кең болса, онда қолданушыға барлық бағандар мен жолдарды олардың атын көрмей қарауға тура келеді. Екінші тәсіл ол, кестелерді жәй қолмен, керекті жолдар мен бағандар санын орналастырып оларға мәліметтерді басу.

Үшіншісі, кестені СОТ арқылы өткізуге болады. Бұл екінші тәсілге қарағанда көп уақытыңызды үнемдейді, бірақта потенциалды қателер саны өседі. Кейбір бағандар кейде бірігіп кетеді және кейде программа нүкте мен үтірді танымайды.

Бейнелер.

Құжаттарда бейнелердің негізгі үш түрі бар.

Ақ-қара суреттер

Ақ-қара фотолар

Түрлі-түсті фотолар

Ақ-қара суреттерді сканирлеуді штрихты режимде (line art mode) жүргізіп және GIF немесе PNG файлы ретінде сақтаған жөн. Ақ-қара фотосуреттерді greyscale режимінде сканирлеп, GIF немесе JPEG файлы ретінде сақтау керек. Түрлі түсті фотоларды сәйкесінше түрлі-түсті сканерде түрлі-түсті режимде сканирлеп JPEG файлы ретінде сақтау қажет. Әдетте, орташа сапалы JPEG файлында қажетті анықтық шешімі бар.

14 СОТ: СИМВОЛДАРДЫ ОПТИКАЛЫҚ ТАҢУ

Қатты дискіде немесе CD-ROM-да сақталған коллекциядағы суреттердің көбі кеңістікті көп қажет етеді. Сондықтан бейне орнатылымдарын оны мүмкіндігінше анық және түсінікті етіп, сонымен бірге оның өлшемін кішірейтіп оптимизациялау өте маңызды. Орынды үнемдеу үшін мәтінге қатысы жоқ кейбір бейнелерді енгізбеген дұрыс.

Суреттер жеке-жеке сканирленуі қажет. Біз бұл суреттерді құжат атының бірінші 6 әрпімен, ал қалған цифрлармен ол орналасқан бет нөмірлерін белгілеуді ұсынамыз. Балама ретінде, әрбір құжат өз директориясында орналасқан деп есептеп, әріптен кейін осы бейне орналасқан бетті қойып пайдаланса болады. Егер бір бетте бірнеше суреттер болса, онда ол u7548e36.jpg немесе p36/jpg деп аталған файлға орналастырылады. Бейнелер сканирленгеннен кейін, сіз суреттің өлшемімен орналасуын түзету үшін арнайы программаларды қолдана аласыз.

Арнайы материал

Құжаттардың көбінде арнайы символдар, формалар және анық емес беттер сияқты техникалық терминдер бар. Бұл қиын танылатын символдар, әдетте әртүрлі тілдермен байланысты. Әр құжат үшін сізге СОТ программасының опцияларынан сәйкес тілді таңдауыңыз қажет. Формулаларды қолмен қайта теру керек, себебі көп жағдайда СОТ оларды танымайды, сондықтан оларды Word-қа енгізу қажет. Анық емес беттерде қандайда бір бүлінуден т.б. түсінуге қиын материалдар болуы мүмкін, сондықтан мұндай жерлерді қайта теріп енгізу қажет.

3.2. Өнімділік пен қол жетерлік ресурстар

Бұрын ескертілгендей СОТ процесінің күрделілігін жете бағаламауға болмайды. СОТ процесін, сканирлеуден бөлек қарастыру керек болса да, оны бағалау үшін ұқсас практикалық басшылықтар пайдаланылады: компьютерлерді сатып алуға қажет ресурстар, жұмыс күшінің болуы мен басқара білуі; зарплата; өндеуді қажет ететін беттердің жалпы саны және бұл құжаттарды үшінші бір адамдарға беруге болама.

Келесі тарауларда біз сіздермен өзіміздің СОТ программасымен Бельгия, Индия және Румыния сияқты елдерде жинаған тәжірибемізбен бөлісеміз. Барлық зерттеулер, санаулар мен цифрлар келесі типтес жағдайлар: кітапханалар мен архивтердің көпшілігінде кездесетін күрделілігі орташа құжаттар (кестелер мен бейнелерді қоса), нәтижелердің жоғары сапалылығы, және жұмыстың ұзақтығы орташа-ұзақ үшін жасалған.

Интенсивті СОТ

СОТ қиын процесс. Ол үлкен жинақылық пен біліктілікті қажет етеді. Ең жоғары нәтижеге жетуге ұмтылудан бұрын 6 аптадай жұмыс істеу керек, бұл процесс кезінде тәжірибеніз өседі.

Ең жоғары өнімділікке жұмыс күнінің алғашқы сағаттарында жеткізіледі. СОТ пен үш сағат жұмыс істегеннен кейін өнімділік күрт төмендейді, шамамен бастапқы деңгейден 50%-ға, ал 6 сағат жұмыстан кейін адамдардың көпшілігі қатты шаршайды.

Алғашқы апталарда дәл осындай жағдайлар болады. Алғашқы апталарда өнімділік жоғары деңгейде болып, содан кейін адамдардың 2/3 шаршап жұмысқа қызығушылығын жоғалтады. Мұндай адамдар кетеді, немесе жұмыстың сапасы мен өнімділігіне әсер ететін өте төменгі деңгейде жұмыс істеуді жалғастырады. Қиын деген 1-5 аптаға шыдағандардың өздері де жұмыс командасының бір бөлігі бола тұра, көбіне 6-12 ай аралығында жақсы жұмыс іздеп, кетеді.

3.1.-бөлімінде жұмыс күшіне тиісті жасалған ескертпелер әсіресе интенсивті СОТ-қа қатысты. Сапасын қадағалауды, құжат жазылған тіл ана тілі болатын немесе ол тілді жетік білетін адамдарға тапсырған жөн. Ең жақсы жұмысшылар, ол оқушылар мен студенттер. Себебі жұмыстың мұндай түрінде үлкен буын мен егде адамдарға қарағанда жастар зейіндірек және жинақырақ. Таңдаудың шартты критеріі мынадай – 25 жастан үлкендерге қарағанда 18-23 жастағы адамдар мұндай жұмысқа жақсы келеді. СОТ процессі ұсақ-түйек жұмысы көп және зеріктіретін жұмыс. Сондықтан қалайда жұмыс рухын көтеру қажет.

Жоғарыда айтылған жәйттар қорытындысы мыналар:

- Жұмысқа ең жақсы кандидаттар 18-25 жастағылар
- Жұмыстың алғашқы сағаттары ең өнімді болғандықтан, жұмысшыларды жарты күнге жалдап тек ең ынталы жұмысшыларға толық жұмыс күн істеуге рұқсат беруге болады.
- Жұмысшылардың үштен екісі шаршағандықтан немесе 3-5 апта бойы қыруар жұмыстан жалыққан, бұның бәрі жұмыстың өнімділігіне әсер етеді. Үйретуге кететін шығынды жабу үшін, жұмысшылардың ұдайы келіп тұруы керек, жұмыс рухын қолдап отыру қажет.

16 СОТ: СИМВОЛДАРДЫ ОПТИКАЛЫҚ ТАҢУ

Кесте 2 СОТ процесінің өнімділігі

	Жұмыс сағ/күн	Бет/күн	Бет/ай
Бастапқы дайындық (6 апта)	3	6	120
Өнімділіктің оптимальды деңгейі	3 7	9 28	150 дейін 200 500 дейін 600

2-кестеде СОТ-да жұмыс істеудің өнімділігінің орташа цифрлары келтірілген. Құжаттар әртүрлі өлшемде және мазмұнда келеді, осы кестені түзегенде бейнелермен кестелердің орташа саны бар құжаттар әртүрлілігінің келесі факторлары ескерілген: мысалы әрбір сегіз бетке бір суретпен 5-те 5-тік бір кесте.

Бейнелер орта-жоғары сапалы деп ойланған, бұл бұрын қарастырылған, ол сканирлеудің сапасына және жұмысшылардың құжат жазылған тілді жақсы білуіне байланысты.

Кестеде сондай-ақ үйрену үрдісінде жүргендермен оптимальды режимде жұмыс істеушілерге арналған мәліметтер берілген. Егер басқарушы команда мүшесі СОТ үрдісіне күніне 3 сағатқа дейін бөлсе, ол бір айда 180-200 бет көрсеткішіне жете алады. Жақсы дайындықтан өткен, жоғары зейінді істі білетін тұрақты жұмысшыға айына 500-600 бетке жету сөз емес.

Сонымен қатар, сапасы нашар түсініксіз кестелер мен суреттері көп құжаттар үшін бұл цифрлар төмен – тұрақты жұмысшы үшін 300-400 бет болады.

Дәлелденген тұрақты жұмысшының айлығы айына 400\$ деп алып көріңіз, ал менеджмент, компьютерлерге, жалға, коммунальдық қызметтер және т.б. шығын 1 адам/айға 300-400\$. Содан кейін 1 бет СОТ-қа 1.2-1.6\$.

Дайындық кезеңін, жалпы көлемін, уақытын, бизнес жабылғанда жұмыстан шығару шығындарын ескерсек бұл цифрлар 1 бетке 1.5-2.5\$-ға дейін көтеріледі.

Жұмысты орындау бағасын өздеріңіз бұл жұмысты кәсіби СОТ компанияларына тапсырғандағы бағамен салыстыруыңыз керек. Румыниядағы кәсіби емес ұйымдар осындай қызмет көрсетеді, және олардың гуманитарлық коммерциялық емес ұйымдарға арналған бағалары төмен және

бір бетке 1.2\$-дан 2\$-ға дейінгі аралықта болады. Егер сіздерде қандайда бір сұрақ туындаса бізге Scanning@humaninfo.org адресіне жазыңыз.

3.3. COT-дың альтернативасы

COT-дың екі альтернативасы бар.

Қолмен қайта теру.

Бұл тәсілді қолданғанда, қарапайым мәтіндік редакторды пайдаланып сканерді мүлде қолданбауға болады.

Бұл жұмысты орындайтын адамдардың, мәтіннің мәнін түсінуі міндетті емес. Олардан тек өздері көріп тұрғанды, қатесіз, ұқыпты теру талап етіледі. Дегенмен, қайта теруде қателер жіберіледі, сондықтан double-keying (қосарлап теру) тәсілі қолданылады. Екі адам бір мәтінді тергеннен кейін арнайы программа осы құжаттың электрондық жасалымының екеуінде сөзбе сөз түп нұсқасымен салыстырып тексереді. Екі жағдайда да бірдей жазылған сөз дұрыс деп есептеледі, бірақ әр кезде бұлай бола бермейді, сондықтан үш жақты теру тәсілі де практикада қолданылады.

Қайта теру тәсілін қолданудың оңтайлы жақтары, ол шығындарды азайтатындығы, себебі COT программалары мен оған қажетті қымбат компьютерлерді сатып алудың қажеті жоқ, ескі модификациядағы компьютерлерді немесе қолдануда болғандарын пайдалануға болатындығы. Сондай-ақ жұмысты мамандығы төмен жұмысшылардың орындай беретіндігі. Бір кемшілігі, ол ең кемінде екі айдай дайындық кезеңінің керектігі. Бір адам тергенде көптеген қате жібереді, сондықтан екі жақты немесе үш жақты теру жүргізу, қосымша шығындармен байланысты.

Барлық шығындар айлық деңгейіне байланысты. Әдетте дамушы елдерде терумен айналысатын адамдар айына 150\$ алады. Олардың өнімділігі күніне 20-30 бет болуы мүмкін. Сәйкесінше айына 400 бет суреттермен бірге. Екі жақты теруде бұл айына 300\$ және басқада шығындарды құрайды.

Бейнелер файлы

COT-ға ең арзан альтернатива, ол барлық сканирленген құжаттар үшін жәй PDF форматын қолдану. Оның бағасы COT бағасының бір ғана бөлігін – 1 бетке 0.1\$ ғана құрайды.

Сканирлеу аяқталып, TIFF файлдарына қолыңыз жеткеннен кейін, автоматты конвертер (әдетте Adobe Acrobat, Adobe Photoshop) TIFF форматындағы барлық файлдарды PDF файлына түрлендіре алады.

Мұндай құжаттардың теріс жақтары, олар бойынша іздеу жүргізуге болмайтындығы. Сонымен қатар олардың көлемі өте үлкен, әдетте бір бетке 50 кб және TIFF файлының сапасына байланысты) болады.

PDF файлдары тек қана СОР-қа ақшаңыз болмаса және жылдам Интернет байланысы бар аздаған адамдардың қолдануына арналған құжаттар үшін ғана қолдануға болады.

3.4. Сканирлеуді СОР-мен бірлестіру

Егер сканер тікелей СОР программасы орнатылған компьютерге қосылса, онда бұл программалардың көпшілігі сканирлеу мен СОР үрдісін бір уақытта жүргізе алады. Бұл тәсіл, егер сіз көлемі шағын ақпараттармен істесеңіз жақсы стратегия болып табылады, ал оның көлемі үлкен болса, өте көп уақытты алады.

Егер сіз айына 100-150 беттік қарқынды ұстанатын болсаңыз, онда бұл тәсіл сізге арналған. Көлемдері үлкен құжаттар үшін сканирлеуді, СОР тан бөлек жасаған тез және ыңғайлы болады.



Үш мысал:

1000-нан 100.000 бетке дейін

4.1 Шағын коллекция: 500 беттен 1000 бетке дейін

ҒӨБ көпшілігінде 5000-10000 бетке дейінгі материалдарды сканирлеу қажет. Мұны дайындығы жоғары жұмысшылар болса еш қиындықсыз COT көмегімен жасап жүзеге асыруға болады.

Сканирлеу

Алғашқы қадам - әр бетке бір файлдан, сапалы электронды TIFF форматта алу үшін материалды сканирлеу, күйге келтірулерде сәйкесінше суреттің әр типіне сәйкес келетін режимді таңдау. Міндет 1000 бетті сканирлеу керек дейік. Бұны жұмысшыларды 1 айға тек сканирлеу үшін жарты ставкаға жалдап ұйымдастыруға болады. TIFF файлдары қатты дискідегі 60-тан 80 Mb кеңістікті алады, ең дұрысы бұл файлдарды компакт диск CD-ROM-ға жазу керек. Жұмыс істеу үшін 100-300\$-лық қымбат емес столға қоятын сканер жеткілікті. Сканирлеуді кешке немесе демалыс күндері жүргізуге болады.

COT

Келесі қадам – ол тілді жақсы білетін бір немесе бірнеше жұмысшының COT процесін жүргізуі. TIFF файлдарын бірнеше компьютерге бөлуге немесе барлық жұмысты бірақ компьютерде жүргізуге болады. Әдетте бұл жарты ставкадағы (мысалы аптасына 20 сағаттан) жұмысшының, толық және қатесіз 1000 бетті Word файлдарына немесе HTML құжаттарына ауыстырып өңдеуі үшін шамамен бес немесе алты ай уақыт алады.

Басқаларға тапсыру

Балама ретінде бұл жұмысты кәсіби компанияға тапсыруыңызға болады. Word және HTML файлдарын алу үшін, бұның бағасы шамамен 1500-2000\$ болады.

20 ҮШ МЫСАЛ: 1000-НАН 100.000 БЕТКЕ ДЕЙІН

4.2. Мекеменің барлық басылымдары: 5000 бет

Үлкен ұйымдардың көпшілігінің архивтері шамамен 5000 беттік басылымнан, қазір басылмайтын кітаптар мен әртүрлі әдебиеттерден тұрады.

Сканирлеу

Жәй столға қоятын сканер үшін бұл өте көлемді жұмыс. Сканирлеуді (5000 бетке шамамен 400\$) компанияға тапсырған дұрыс. Немесе қағазды автоматты түрде беретін сканер (900\$ мванында) сатып алу қажет. Балама ретінде, сканерді басқа FӨБ пен немесе басқа ұйымдармен бірлесіп (6000\$-ды бірлесушілер санына бөліп) сатып алуға болады. 5000 беттің бәрі шамамен 300-400 Мб кеңістікті алады, сондықтан біз сізге оны қатарласа CD-ROM-ға жазуды ұсынамыз.

СОТ

Алдыңғы жағдайдағыға ұқсас жұмыс күшін пайдаланып сізге қайтадан СОТ үрдісінен өту қажет. Бір немесе бірнеше компьютерді пайдаланса болады. Бұл 5000 бетке аптасына 20 сағаттан істегенде 25-30 ай уақытты алады. Практикада бұл өте ұзақ мерзім. Жұмысты сапалы және қысқа мерзім ішінде аяқтау үшін жұмысшыларға ақы төлеуге, сапалы қадағалауға қажетті орын беруге т.т. жауапты адам керек. Қатты дискінің 300-400 Мб кеңістігін алып, Интернет арқылы өте ұзақ жүктелетін PDF файлын жасауға болады.

Басқаларға тапсыру

Сканирлеу мен қатар СОТ-ды да басқа компанияға тапсыруға болады. Бұл жағдайда барлық жұмыстың бағасы 7500\$-дан 10000\$-ға дейін болады.

4.3 Шағын кітапхана: 100.000 бет

Әртүрлі үлкен ұйымдар, университеттер үкімет және мамандандырылған кітапханалар бүкіл кітапханасын – 100000 беттегі цифрлы форматқа ауыстырғысы келуі мүмкін. Ең бірінші ойланатын нәрсе – барлық басылымдарға деген авторлық құқық. Сіз сканирлеудің алдында, осы құқықты иемденген адамдардың рұқсатын алуыңыз қажет.

ҮШ МЫСАЛ: 1000-НАН 100.000 БЕТКЕ ДЕЙІН 21

Сондай ақ сіз бұл құжаттардың цифрлы форматтарының әлі жоқ екендігіне көз жеткізіп алуыңыз керек.

Сканирлеу

Бұл көлем қағазды автоматы түрде беретін сканер үшін өте үлкен. Бұны біреуге (100000 бетке 8000\$-ға) тапсыру, немесе басқалармен бірлесіп қымбат сканер (6000\$) сатып алу қажет. 100000 бет 6-8 GB кеңістікті алады. Сізге қайтадан көшірмелерді бірнеше компакт дискіге орналастырған дұрыс.

COT

Екінші қадам COT үрдісін (немесе сирек қолданылатын құжаттар үшін PDF файлдарын жасау) жүргізу. Бұл барлық COT үрдісі арқылы Word пен HTML-ге көшіруге жарты ставкадағы жұмысшыларды пайдаланған жағдайда шамамен 500-ден 700 айға дейін алады. Сондықтанда бұны кәсіби шеберлерге тапсырған дұрыс. Шығындарды азайту үшін, сирек қолданылатын беттердің көпшілігі – айталық 80% немесе 80000 бет PDF файлына көшіруге болады, ал қалған 20000-ын Word пен HTML-ға. PDF файлдар Интернет арқылы ұзақ жүктеліп 4-6 GB орын алғанымен оның бағасы кәсіби компанияда бір бетке 0.2\$ (80000 бет 16000 доллар тұрады). Егер TIFF –ты PDF-қа конвертациялауға жалдамалы еңбекті қолданса онда ол жарты ставкадағы адамның 10-20 ай жұмыс уақытын алар еді.

Басқаларға тапсыру

Егер 80% ды PDF та ал 20% HTML, Word та сақтаса, біріншісі 16000\$, ал екіншісі 30000\$-дан -40000\$ дейін тұрады, жалпы 50000 дай доллар тұрады. Егер бәрін COT арқылы өткізсе 150000-200000 доллар болады.

22 ҮШ МЫСАЛ: 1000-НАН 100.000 БЕТКЕ ДЕЙІН



Электронды коллекция жасау

Электронды коллекцияны жасау шешімін қабылдағанда үш басты нәрсені ескерген жөн. Бірінші, коллекция ұйымдастырылған болып, белгілі құрылымы болуы қажет. Мазмұны үлкен болған сайын, оның индекстері түрлі болуы және күшті іздеу жүйесі қажет. Көлемі 3000-5000 беттік коллекциялар үшін жоғарыда сипатталған талаптардың қажеттілігін ең алдымен ескеру керек. Бұл коллекция арналып жасалған адамдар тобын дәл анықтап, оларға кеңес беру қызметін көрсету қажет.

Қаражаттың болуы әлуатті жұмыс көлемін анықтайды.

5.1. Коллекция жасаудың тәсілдері

Интернет беттер негізінде жасалған компакт дискілердің тамаша мысалдары көп. HTML, PDF, Word файлдары бір бірімен гиперсілтемелер арқылы байланысқан. Коллекция ішінде қозғалу әртүрлі сілтемелер, кестелер, түйінді сөздер, индекстер т.т. арқылы жеңілдетілген (басқа сөзбен айтқанда сіз бұл дискіде, кез келген Интернет бетіндегідей қозғала аласыз). Мұндай жүйелер, белгілі бір шекке дейін (бірнеше мың бет) дұрыс жұмыс істейді, бірақ 3000 нан 5000 ға дейінгі беттерде жақсы құрылымдалған коллекция мен күшті іздеу жүйесінің болғаны өте маңызды. Міне осындайда Greenstone-сыз іс тынбайды.

Greenstone программалық топтамасы алдыңғы қатарлы іздеу жүйесі бар құрылымдалған цифрлы кітапхана жасайды. Бір CD-ROM-ға 15000-ға дейін индекстелген беттерді сидыруға болады. Әр CD-ROM Internet сервер қызметін атқара алады.

Сәйкес нұсқаулар коллекция жасау үрдісін сипаттайды. Қарапайым коллекцияларды қолданушыға арналған нұсқауда (3-тарау) сипатталған коллекцияларды пайдаланып құруға болады.

Осында сипатталған күрделірек коллекцияларды жасау үшін ең дұрысы жасаушыға арналған басшылықты (1-тарау) сипатталған командалық жолды қолдану болып табылады. Күрделі коллекциялар жасауда Greenstone-нің барлық мүмкіндіктерін толық меңгеру үшін сізге 2 тарауды оқып шығу қажет.

5.2. Органайзер

Greenstone гуманитарлық мақсаттағы құрылымдары мен ұйымдастырылулары ұқсас бірақ мазмұндары әртүрлі бірнеше коллекция жасау үшін қолданылған. “Development Library Subset” (DLS) және Демо-коллекциялары осындай құрылымға негізделген.

Органайзер программасы сондай-ақ Greenstone программалық топтама құрамына кіріп, DLS –тікі сияқты құрылымы мен интерфейсі бар коллекция жасау үрдісінде қолданылады. Ол командалық жолды (command-line building process) қолданумен коллекция жасау үрдісін бірталай жеңілдетеді. Органайзер Microsoft Windows үшін C++ те жасалған өнім, сондықтан ол Windows-та (95, NT және басқалар) ғана қолданыла алады.

Органайзер цифрлы коллекцияларды ұйымдастыру, құрылымдауды жан жақты жеңілдету мақсатында жобаланған: құжат атауларын енгізу, метамәліметтерді жасау және оларды өзгерту т.т. Органайзер негізгі Greenstone программалық топтамасымен, негізгі программа коллекция жасағанда қолданатын алты аралық файлдар жасау арқылы қарым-қатынас жасайды: collect.cfg, metadata.xml, sub.txt, org.txt, keyword.txt, AZList.txt. Бұл файлдардың құрылымы мен мазмұны қолданушыға арналған басшылықта қарастырылған.

RecPlug плагині Metadata.xml файлын оқу үшін қолданылып жасаушыға арналған басшылықтың 2.1 бөлімінің соңында қарастырылған: ол 10-суретте көрсетілген. Sub.txt, org.txt, keyword.txt, AZList.txt файлдары 2.2 бөлімінде сипатталған иерархиялық классификатор (Hierarchy classifier) қолданатын иерархиялық құрылымды анықтайды: Sub.txt-ның үзіндісі қолданушыға арналған басшылықтағы 15-суретте көрсетілген. Бұл бес файл DLS коллекциясымен (және басқа құсас коллекциялармен) коллекция конфигурациялық файлы (collect.cfg) арқылы байланысқан, себебі metadata.xml файлы конфигурациялық файлда көрсетілген (Rec Plug) көмегімен оқылады, ал қалған төртеуі өзі оқылады, себебі олар иерархиялық классификаторларды анықтайтын конфигурациялық жолдарда анық көрсетілген.

Толық библиографиялық функциялы және толық функциональды рубрикалары бар коллекция жасаймын деушілер осы файлдарды жасауда қиналады. Органайзердің мақсаты – осы файлдарды жасауда көмектесу, және де шынында ол барлық қажетті командалық жолдарды сіз үшін жазады. Дегенмен, қазіргі кезде органайзер төменде сипатталған

тек белгілі бір құжаттар моделімен жұмыс істеуге қабілетті, сондықтан да бұл модельге еңбеген метамәліметтерді қосам деген қолданушылар, қажетті мәліметтерді Органайзер көмегімен жасалған файлдарға қосулары керек. Органайзердің бұдан кеңірек функциясы Greenstone-ге кейіннен қосылады.

Органайзерді орнату және қолдану

Стандарты Greenstone программасын инсталляциялау аяқталған соң, органайзерді орнатуды ұсынатын Терезе пайда болады: Егер «Иә»-ні бассаныз инсталляция аяқталғанша жүреді. Егер сіз «Жоқ» деп жауап берсеңіз, немесе Органайзерді алып тастап, кейіннен қайта орнатқыңыз келсе, онда Greenstone дискісіндегі Windows—Utilities директориіне еніп, organizer.exe көмегімен инсталляциялауды жаңартыңыз.

Программаны Пуск терезесіндегі Programs тақырыбы астындағы Greenstone Digital Library ден төмен орналасқан Organizer-ге басып, іске қосыңыз. Сізден мәліметтер базасын көрсетуді сұрайды: dls.mdb файлын таңдап DLS мәліметтер базасын көрсетіңіз. Содан кейін сізден қолданушы аты мен парольді енгізуді сұрайды: admin сөзін екі терезеге де енгізіңіз.

Органайзердегі мәліметтер базасының әрқайсысында бір немесе одан да көп коллекциялар болуы мүмкін. Органайзерді алғаш қосқанда сіздерде тек бір коллекциядан тұратын, коллекциялар тізімі пайда болады: Development Library Subset (DLS) (Даму кітапханасынан алынған бөлім). Әдетте Органайзермен, жаңа мәліметтер базасын жасаудан (Органайзердің бұл жасалымы мүлде жаңа және бос мәліметтер базасын жасау функциясын қолдамайды) гөрі о баста орнатылған мәліметтер базасына жаңа коллекциялар қосып, алдыңғы коллекциялардың бөлімдерін және өзгерістерін жазып жұмыс істеу оңайырақ.

Мәліметтер базаларының әрқайсысы мен коллекцияда түзетілетін үш тізім болады:

- Коллекцияға енгізуге болатын құжаттар
- Ұйымдар: құжаттармен байланысты ұйым атаулары
- Рубрикалар: рубрикалардың иерархиялық құрылымдарын жасау үшін қолданылады.

Библиографияның қосуға болатын басқада атрибуттары төменде сипатталған. Есіңізде болсын, Органайзерге енгізілген барлық өзгерістер, одан шыққан кезде сақтауды растауды сұрамай ақ автоматы түрде жазылады. Сондықтан да егер сіз мәліметтер базасының бастапқы жағдайдағы артық көшірмесін сақтап алып қойғыңыз келсе сіз оны Органайзерге кірген бойда жасауыңыз керек (артық көшірме жасаудың екі функциясы төменде сипатталған). Жеке коллекцияның артық көшірмесін сақтау

мүмкін емес.

Құжат моделі

Органайзерде, әр құжатқа келесі категориялар бойынша метамәліметтер қыстырылуы мүмкін.

i. Атауы

ii. Ұйым (әдетте құжаттың шығарушысын көрсету үшін, сондай ақ авторға, немесе осы құжатпен қалай да байланысты басқа ұйымға қолданылатын қайталанбалы атрибут)

iii. Рубрика Ұиерархиялық классификациялау үрдісі

iv. Түйінді сөздер (индекстеудің қосымша параметрлері – DLS коллекциясында құжаттарды іздеу үшін қолданады, олардың негізінде белгілі бір тәсіл, мысалы «көбелектерді қалай өндіру» т.т. жатыр).

v. Периодикалық басылым.

vi. Басылым датасы

vii. Бет сандары

viii. Жұмыс нөмірі (Құжаттарды білдіреді, Greenstone-де метамәліметтерді құжаттың мәтінімен және суреттермен коллекцияны жасау кезеңінде байланыстыру үшін қолданады)

ix. Тіл

x. Блок аты (Бірге түрлендірілген құжаттар тобын белгілеу)

xi. Ұсынылатын коллекциялар (жекелеген құжатты потенциалды түрде қосуға болатын коллекциялар)

xii. Авторлық құқықтар туралы белгілер

xiii. Авторлық құқықтар кодексі

Алғашқы бес пункт ерекше, себебі Органайзер Greenstone-ні құжаттарды іздеуді жүзеге асырғанда осы мәліметке негізделген индекстер жасау үшін күйіне келтіреді. Vi және viii атрибуттары индекстелмеген және олар іздеу параметріне кірмейді, бірақ құжатты қосымша белгілеу үшін қолданылады. Ix және xii атрибуттары Органайзердегі құжаттармен тек программаның ішінде жұмыс істеу үшін ғана қолданылады (internal document management). Олардың жасалған коллекцияға ешқандай әсері жоқ.

“Түйінді сөздер” атрибутын индекстеудің қосымша параметрі ретінде қарастыруға болады. Мысалы, ол авторларды немесе бұл құжатты шығарушы елді, жәйғана қажетті ақпараттарды түйінді сөздер метамәліметтеріне орналастырып іздеу үшін қолдныла алады.

Ал қарауға арналған «тәсілдер» кнопкасы “тәсілдер” сөзін “авторлар” немесе “елдер” сөзіне Органайзер жасаған collect.cfg файлының бесінші жолын өзгертіп күйіне келтіре алады.

Greenstone әр құжаттың тараулары мен бөлімдеріне сілтемесі бар мазмұн жасай алады. Қазіргі уақытта бұл функция жұмыс істемейді, бірақ оны осы тараудың 5.3 бөлімінде сипатталған жолмен белсенділендіруге болады.

Құжаттардың әртүрлі моделдерімен жұмыс істеуде Greenstone-ның өте икемді жүйе екендігін есте ұстаған жөн, бірақ қазіргі уақытта метамәліметтер функциясын белгілегенде әр модель, жекелей күйіне келтірілуі керек, олар жасаушыға арналған басшылықтың 2.2 бөлімінде сипатталған коллекциялық файлдарға енгізілуі қажет. Егер қажетті модель DLS моделіне ұқсас болса, онда оны Органайзер жасаған (build files) жасау файлдарын қолданып Greenstone-ге туралап көруге болады, немесе оларды Greenstone ұсынған басшылықтарға сәйкес қолмен күйіне келтіруге болады.

Органайзерге шолу

Міне Органайзердің әртүрлі панельдеріне шолу:

i. (Органайзердің бас терезесі)

Бұл терезе мәліметтер базасын таңдағаннан кейін пайда болады. Ол төрт бөліктен тұрады:

- (мәзірдің көлденең панелі) төменірек орналасқан көлденең жарлық панелімен қоса.
- Вертикалді панель сол жағында
- Орталық аумақ оған вертикаль панельден таңдалған барлық ақпарат шығарылады.

A. Мәзірдің көлденең панелі әртүрлі сегіз мәзірден тұрады, олардың әрбіреуі бір не оданда көп командалардан тұрады:

- Файл (file) мәзірінде тек бір команда бар: шығу (exit) Органайзерден шығу үшін.
- Жасау (New) мәзірі коллекцияға құжаттарды, ұйымдарды, рубрикаларды қосу немесе өзгерту және list of all

- possible attributes (мүмкін болатын барлық басқа атрибуттар тізімін) құжат моделіне қосып, өзгерту мүмкіндігін береді.

New/Organization командасы Edit Organization Name терезесін шығарады (ұйым атауларын түзету) мұнда ұйымның толық және аббревиатураны (егер Auto complete) автотолтыру таңдалса онда бас әріптен жазылған сөздердің бірінші әріптері аббревиатура ретінде сақталады. Егер Авто толтыру функциясы сәйкес келетін аббревиатураны ұсына алмаса (бас әріптердің жоқ болуына байланысты немесе мұндай аббревиатура мәліметтер базасында болса), онда бұл функцияны алып тастап, аббревиатураны қолмен енгізу қажет.

New/Subject командасы Add new subject (жаңа рубриканы қосу) терезесін шығарады, ол жаңа рубриканы мәліметтер базасындағы жалпы (иерархиялық емес) рубрикалар тізіміне (global subject list) қосады. Бұл терезе сондай ақ түйінді сөздерді (KeyWords) осы тізімде жасауға пайдаланыла алады.

- Delete (алып тастау) мәзірі коллекцияларды, құжаттарды, ұйым аттарын, рубрикалар мен құжат моделіндегі басқа атрибуттардың мүмкін болатын мәндер тізімінен заттарды алып тастауға мүмкіндік береді.

- Edit (түзету) мәзірінде Copy (көшіру) (Сондай-ақ Ctrl және C пернелерін бірге басуға болады) командасы бар оның көмегімен көшіру функциясы орындалып, алмасу буферінде сақталады кейіннен бұларды Органайзердің басқа элементтеріне қыстыруға болады.

- View (қарау) мәзірі экранға орталық аумақта шығарылатын коллекциялар тізімін, құжаттарды, ұйымдармен Рубрикаларды шығару командаларын орындау мүмкіндігін береді. Бұл терезеде сондай ақ жарлықтық панельдерді көрсетуді белсенділендіруге болады.

- Tools (Қызмет) мәзірінде:

Documents командасы онда екі функция бар: біріншісі ұйымның атын басқаға құжаттардың “ұйым” атрибутында

(ең алдымен ұйымның атын, ұйымдардың жалпы тізіміне қосу қажет) толық өзгерту; ал екінші функция **құжаттардың жалпы тізімінен (Documents-global-list) таңдалған** құжаттарға бірнеше атрибут беру.

Collections (коллекция) командасында екі құрал бар: Statistics (статистика) құралы коллекция жайындағы қажетті статистикалық мәліметтерді көрсетуге арналған, ал Export lists (тізімдерді экспорттау) құралы мәтінге атаулар тізімін немесе коллекцияның үш басты атрибуттарының біреуін экспорттау үшін қолданылады, тізімдер стандартты аттармен \Program Files\Human Info\Organizer\Data Base\ директориінде сақталады. Backup (көшірмелерді сақтау) командасы бүкіл мәліметтер базасының сақтандыру көшірмесін жасау үшін қолданылады, әркез сіз бұл команданы орындағанда (мәліметтер базасы стандартты атпен, сақталған датасы және уақыты көрсетіліп \Program Files\Human Info\Organizer\Data Base\ директориінде сақталады)ю Бір мәліметтер базасының бірнеше көшірмесін жасауды болдырмас үшін сіздің растауыңызды сұрайды.

- Administration tools (Администрациялау құралдары) мәзірі жаңа қолданушыларды қосу мен алып тастау, қолданушының паролін өзгерту мүмкіндігін береді. Тек администратор ғана мәліметтер базасына қолданушының « администратор » немесе «қонақ» ретінде енуіне рұқсат беріп, парольдерді өзгерте алады.
- Help (Көмек) мәзірінің Microsoft-тікіне ұқсас Help topics анықтамалық жүйесі бар, онда About Organizer (Организер туралы) бөлімінде жасалымы мен жасаушылар жайлы ақпарат бар.

Жарлықтық панель көлденең панельден төменірек орналасқан және орталық аумаққа не шығарылғанына байланысты ячейкалар динамикалық түрде өзгереді. Сол жақтағы екінші иконканы басы, орталық аумақтан таңдалып алынған элементті екі рет басқан мен парапар, сонымен бұл элементті (документ, коллекция болсын) түзетуге болады, соңғы иконканың алдындағы (A-Z әріптерімен бейнеленген) иконка сіздерге орталық аумақтағы элементтерді алфавит, қосылған датасы және т.т. бойынша сорттауға мүмкіндік береді. Бірінші, үшінші және соңғы иконкалар жоғарыда сипатталған үш команданың функциясындай функцияны орындайды:

Edit/Copy; көрсетілген бетте жаңа элемент жасау; New/Collection/Empty, New/Document, New/Organization немесе New/Subject; және Help/About Organizer.

B. Vertical toolbar (Вертикальды панель) бес кнопка бар. Бұл панельдегі төрт кнопка келесі элементтерді түзету үшін қолданылады: коллекциялар, құжаттар, ұйымдар, рубрикалар (бұл тізімді көлденең панельдегі View мәзіріндегі сәйкес команданы таңдап алуға болады):

- Коллекция тізімі: Collections (коллекциялар) (Ол үнсіздік бойынша мәліметтер базасына енген кезде тұрады) кнопкасын басқаннан кейін, мәліметтер базасындағы барлық коллекциялар тізімін алуға болады. Collection Properties (коллекция қасиеттері) терезесін ашу үшін коллекцияның кез келген атын екі рет шертіңіз онда сіз бұл коллекцияға жаңа құжаттар қоса аласыз немесе осы коллекция құжаттарының атрибуттарын қосып/түзете аласыз. Жаңасын жасау үшін жоғарғы панельде немесе жарлықтар панеліндегі сол жақтан үшінші иконка (New/Collection, жасау/коллекция) командасын қолданыңыз.

- Құжаттар тізімі: Documents (құжаттар) кнопкасын таңдағанда мәліметтер базасында бар барлық құжаттардың тізімі (коллекцияда бары да, және әлі жоғыда) пайда болады. Құжаттардың атрибуттарын қарап, өзгерту үшін құжат атын екі рет басыңыз. Жаңа құжат жасау үшін Create/Document (жасау/құжат) командасын орындау керек ол үшін жоғарғы панельдегі немесе жарлықтар панелінің сол жағындағы үшінші кнопканы басыңыз.

Атауы бойынша іздеу: құжаттардың атауы бойынша іздеуді Document title тақырыбының жоғары жағындағы кішкене терезеге қажетті сөзді енгізіп жүргізуге болады. Бірінші іздеудің нәтижелерін қарау үшін “Дүрбі” иконкасын басыңыз. Одан кейінгі нәтижелерді көру үшін оң жағындағы келесі (“көрсеткіші бар дүрбі” иконка) кнопкасын басыңыз.

Құжаттарды филтрлеу: кейде, бірғана тілде, бір ұйым жазған немесе бір рубрикадағы құжаттарды ғана түзеткен оңай. Ол үшін құжаттың жоғарғы оң жақ бөлігіндегі Filter (Филтрлеу) (воронка бейнесі) кнопкасын таңдаңыз, терезені көрсету үшін Search documents (құжаттарды іздеу). Өз критерилеріңізді енгізіп оларды қабылдау үшін Apply filter (филтрді қолдану) кнопкасын басыңыз. Сіз кез келген уақытта филтрді қолдана аласыз немесе оны алып тастай аласыз, ол үшін Apply filter опциясын алыптастайсыз. Бұл механизм Boolean «and” іздеуіне ұқсас: көрсетілген барлық құжаттар осы филтрдің критерііне сәйкес келуі керек.

- *Ұйымдар тізімі (organization list)*: сіз мәліметтер базасындағы барлық ұйымдар атын *Organizations (ұйымдар)* коллекциядағы құжаттарға байланған кнопкасын басып қарай аласыз. *Edit Organization Name (ұйым атын түзету)* терезесін ашу үшін ұйымдардың кез келген атын басыңыз, ұйым атымен оның аббревиатурасын қарап/түзету үшін (жасау/ұйым командасын орындағанда пайда болатын терезе сияқты терезе шақырылады). Ұйымның жаңа атын жасау үшін *New/Organization* командасын қолданыңыз, ол жарлық панелінде жоғарыдағы панельде сол жақтан үшінші иконка (жоғарыға қара).
- *Рубрикалар тізімі (Subjects list)*: Рубрикалар коллекция құжаттарын классификациялаудың стандартты элементі. *Subjects (Рубрикалар)* кнопкасын басу мәліметтер базасындағы рубрикалар тізімінің барлығына, рубрикалар құжаттарға қыстырылмаған болса да қосу, алып тастау, және түзетулер жүргізуге мүмкіндік береді. Кейіннен коллекциялық деңгейде пайдаланып рубрикаларды құрып коллекциялық құжаттарға байлау үшін шексіз көп рубрикалар жасауға болады. *Edit subjects (рубрикаларды түзету)* терезесін ашып рубрикалар қасиетін қарап/түзету үшін тізімдегі кез келген рубриkanı басыңыз. Жаңа рубрика жасау үшін *New/Subject* кнопкасын басыңыз.

Edit Subject (Рубрикаларды түзету) немесе *Add new subject (жаңа рубрикаларды қосу)* терезелерінде, рубрикалардың жалпы тізіміне (*subject global list*) “түйінді сөздерді” қосуға болады. (*New Keyword name терезесі, New/Add-Modify Keywords түйінді сөздерді жасау/қосу-өзгерту*) командасы сияқты істейді, содан кейін коллекцияға қосу үшін іздеу сәтінде рубрикалар категориясын қолдану кезінде белгілі бір рубрикалар қасиеті ретінде (*Select for this subject box (осы Терезе үшін таңдалған көрсеткіштер)*) күйіне келтіріледі. Түйінді сөздер (көрсеткіштер категориясына жатқызылған, жатқызылмаған) кейіннен құжатқа оның атрибуты ретінде қосыла алады («тәсілдер» DLS коллекциясында). Байқадыңызба тізімде алфавиттік тәртіпте түйінді сөздердің екі түрі араласып кеткен, бұл түйінді сөздерді құжаттарға Берген кезде аздаған ыңғайсыздықтар тудыруы мүмкін: белгілі бір көрсеткіштерге берілетін түйінді сөздер алдына «Z-“ кодын қосу арқылы мұны болдырмауға болады.

Зейін қойыңыз: егер басты мәліметтер базасы өте үлкен болса, коллекция элементтерін жүктеу өте ұзақ уақыт (минут немесе оданда көп) алуы мүмкін. Жаңа коллекцияны таңдағаннан кейін, онымен жұмыс істеудің алдында оның барлық элементтерінің толық жүктелуін күтіңіз, сіз жүктеу процесі кезінде лампочка түріндегі иконканың пайда болғанын байқайсыз. Егер сіз толық жүктелуді күтпесеңіз, онда программаның жабылып қалуы мүмкін.

Export Files (файлдарды экспорттау) кнопкасы цифрлы кітапхана құрылымын Greenstone Digital Library программасының өзіне экспорттау мүмкіндігін беретін Export settings (экспорттауды күйіне келтіру төменде сипатталған) терезесін ашады.

i. Коллекция қасиеттері (Collection properties) терезесі.

Бұл Терезе, қолданушыға коллекциялар құру немесе өзгерту мүмкіндігін беріп, Органайзердің басты терезесіндегі коллекциялар тізімінен коллекцияны таңдағаннан кейін пайда болады. Ол қолданушыға коллекцияларды шолудың төрт режимін жоғарғы жағындағы тиісті кнопканы басып таңдау мүмкіндігін береді. Шолудың әрбір режимі, төменде сипатталған мәліметтерді таңдау мен түзету функцияларымен жұмыс істеу мүмкіндігін береді:

a. *Subjects view (Рубрикалар бойынша қарау)*: бұл жерде алты деңгейге дейін тармақталған коллекциялардың көпшілігіне үш деңгейден артық керек болмаса да иерархиялық құрылымды кез келген коллекция үшін саны шектелмеген рубрикалар жасауға болады. Коллекцияға жаңа рубриканы қосу үшін (рубриканы бір немесе бірнеше құжаттардың атрибуттары ретінде қолдану үшін) i) жаңа рубриканы сақтағыңыз келген тақырыпты таңдаңыз ii) add subject (рубриканы қосу) кнопкасын басыңыз iii) Add Subject from Global List (жалпы тізімнен алынған рубриканы қосу) –ды таңдаңыз iv) кнопканың көмегімен пайда болған рубрикалардың жалпы тізімін қараңыз v) қажетті рубриканы таңдаңыз vi) ОК-ді басыңыз. Егер бұл рубрика жалпы тізімде болмаса онда алдымен Add subject командасының терезесінде орналасқан Add New Subject (жаңа рубриканы қосу) опциясын қолданыңыз (басты мәзірге оралып, оны New/Subject командасының көмегімен қосу мен пара-пар). Рубрикаларды сондай-ақ, рубрика мен құжат бейнесінің арасында орналасқан Edit (түзету) иконкасының көмегімен өзгертуге болады (рубрикалар атына жасаған барлық өзгерістеріңіз тек осы коллекцияға қолданылып, рубрикалардың жалпы тізіміне әсер етпейді).

Қандайда бір рубрикаға бір немесе одан да көп құжаттарды қосу үшін (рубриканы бір немесе одан да көп құжаттарға байлау), рубриканы таңдаңыз. Содан кейін тізімнен бір немесе бірнеше құжаттарды таңдаңыз, және «жоғары қаратылған көрсеткіштер мен қызыл кітап» бейнесі бар иконканы басыңыз (немесе қажет құжаттардың әрқайсысын екі рет басыңыз). Сонымен, белгілі бір рубрикаларға бекітілген құжаттар «қанатшамен» белгіленген; ыңғайлылығын қарау үшін сіз оларды экранда «жоғары» және «төмен» иконкаларының көмегімен қозғалтуға болады, олар экрандар арасында орналасқан.

Сіз құжаттардың, рубрикаларының иерархиялық классификациясына қосылғандығын көресіз. Содан кейін сіз бұл амалды қалған құжаттарды сәйкес рубрикаларға табыстау үшін қайталай аласыз. Бір құжатты рубрикалардың кезкелген санына шектеусіз қыстыра аласыз. Құжатты рубрикадан алыптастау үшін, осы рубрикадан құжат атын таңдаңыз да алыптастау (delete)-ні басыңыз.

- à. *Organization's view* (ұйымдар бойынша қарау): қараудың бұл режимі қандайда бір құжаттарды жасауға қатысқан ұйымдар тізімін қарау мүмкіндігін береді. Бұл опция құжаттарды іздеу үшін кезекті критерий ретінде қолданыла алады (құжаттарды қосып және оларға “ұйым” атрибутын құжаттар мәзірінде беруге болады). Сол жақ бағандағы тізімде, коллекциядағы сәйкес құжаттарға бекітілген ұйымдар аттары бар; оң жақ терезедегі тізімде ұйымдар бекітілген құжаттар орналасқан – аттары «қанатша” белгісімен белгіленген құжаттар коллекцияға енгізілген, ал қалғандары әлі енгізілмеген. Коллекциядан алып тастағыңыз/қосқыңыз келген кез келген құжат атына бассаңыз болды. Аттарды таңдау немесе алып тастау үшін төменгі оң жақ бөлігінде орналасқан (қанатшасы бар ақ квадрат немесе бос ақ квадрат) иконкасын қолданыңыздар. Құжаттарға әлі байланбаған ұйымдардан құжаттарды таңдау үшін, сіз коллекция тізіміне жаңа ұйымдар атын қосуға арнақлған) Add Organizations from Global List ұйымдарды жалпы тізімнен таңдау опциясын таңдауыңыз керек.

Сіз коллекциядан таңдалған ұйымға бекітілген барлық құжаттарды алып тастау үшін (тек коллекциядан ғана, ұйым жалпы тізімде қалады) Remove organization (ұйымды алып тастау) кнопкасын қолдана аласыз. Тек қана коллекциядағы құжаттармен жұмыс істеу үшін List only selected organizations опциясын таңдаңыз. Бұл режимде сіз сондай-ақ жаңа ұйымдардың аты мен жаңа құжаттарды жалпы тізімге қоса аласыз (Add Documents мәзіріндегі Add new document немесе Add organization мәзірі, бас мәзірге қайта оралып, оларды New/Organization немесе New/Subjects командасының көмегімен қосқанмен бірдей).

- с. Documents view (құжаттар бойынша қарау): таңдалған коллекцияның барлық құжаттарының тізімі шығарылады. Бұл рубрикалар бойынша қарау командасы төменгі терезесінде пайда болатын тізіммен бірдей. Айырмашылығы құжатқа екі рет басқанда құжат қасиеті терезесі пайда болуында. Бұл терезеде сіз құжат атрибуттарын сәйкес опцияларды таңдау жолы арқылы қосып/өзгерте аласыз (рубрикалар бойынша классификациялау мен байланысқан атрибуттардан басқа атрибуттардың барлығы:

тех, которые связаны с классификацией по рубрикам) путем выбора соответствующих опций:

- General (жалпы мағлұмат): бұл терезеде сіз құжат атын, жұмыс нөмірін, бет санын және суреттер санын, жариялану жылы мен блок атын енгізе аласыз. Сіз Find images (суреттерді іздеу) кнопкасын басып және құжаттың барлық суреттері сақталған директориін таңдап автоматты түрде суреттер санын анықтап, көрсете аласыз. Содан кейін Extenstions (форматтар) алаңында суреттер сақталған форматтардың (GIF, JPEG және т.т.) барлық к-рсетініз.
- Advanced (кеңейтілген режим): Егер құжат мерзімдік журнал немесе қандайда бір басылым сериясының жеке номері болса, онда сіз оныда көрсете аласыз, сонда ол мерзімді басылымдарды іздеудің арнайы опциясын жасайды. Осы терезеде сондай-ақ құжатты ұйыммен байланыстырып, оның қай тілде жазылғандығын көрсете аласыз.

Егер құжат бірнеше ұйымға тиісті болса, немесе бірнеше тілде мысалы Ағылшын/Француз жазылса, онда сіз бұларды сәйкес терезелерде көрсетуіңіз керек.

- Copyright (Авторлық құқық): әр басылымның авторлық құқық статусын білу өте маңызды. Бұл терезе екі бөлікке бөлінген: бір бөлігінде басылымның авторлық құқық статусы орналасқан. Бұл мәлімет құжаттармен Органайзерде жұмыс істеу үшін қажет және ол Greenstone программасының жұмысына әсер етпейді.
 - Keywords (түйінді сөздер): сіз мәтін түріне тән түйінді сөздерді көрсете аласыз. Бұл қасиет іздеумен рубрикаға бөлуде қосымша критерий бола алады. DLS коллекциясында бұл қасиет «тәсілдер» параметрін қолдану үшін пайдаланылады. Бірақ метамәліметтердің кез келген басқа түрі үшін қолданыла алады, мысалы автор немесе шығарушы ел.
- Байқадыңыз ба, Subjects, Organizations, Documents views режимдерінде Add documents кнопкасы қолданушыға жаңа құжаттарды

турасынан диалогтық терезеден немесе мәліметтер базасының жалпы тізімінен (global database list) құжатты таңдау жолымен қосуға мүмкіндік береді. Коллекция қасиеттері терезесіне (Collection properties window) қосылған құжаттар, автоматы түрде осы жалпы тізімге енгізіледі. Коллекцияға жалпы тізімнен жаңа құжат қосылғанда, қолданушыға қажетті құжаттарды белгілі бір критерилер бойынша табу мүмкіндігін беретін Search documents (құжаттарды іздеу) терезесі пайда болады. Бұл баяғы құжаттар тізімін (documents list) талқылағанда сипатталған фильтрация функциясы.

a. Others view (қараудың басқа режимдері): Бұл қарау режимі (өзгертуге болмайтын) мерзімдік басылымдар аттары тізімін және сонымен қатар құжаттардың аттары классификациясын әр тіл бойынша алфавиттік тәртіпте көрсетеді. Қолданушы алфавиттік классификациялық топтарды (мысалы, А-В, Е-К немесе А-Л және т.т.) оларды кітапхана экранында ыңғайлығып орналастыру үшін өзгерте алады. Ол үшін, Split letters (әріптерді бөлу) кнопкасын пайдаланыңыз, сіз олардың арасындағы оптимальды аралықты тапқаннан кейін нәтижені Save кнопкасын басып сақтай аласыз (сақтауға дейін сіз алдыңғы әріптік категорияларға Load/Refresh (жүктеу) басып, және де Eliminate (Алып тастау) кнопкасын басып әріптік категорияларды алып тастай аласыз).

ii. Экспорттау терезесін күйіне келтіру

Бұл функцияға вертикальды панельден қол жеткізіледі және сіздің түпкілікті нәтиженізді кітапхананың өзіне экспорттау мүмкіндігін береді, әдетте бұл коллекция жасауға байланысты соңғы саты.

Export Settings (Экспорттауды күйіне келтіру) терезесін шақыру үшін Export Files-ті таңдаңыз. Экспорттайтын коллекцияны және оны орналастыратын директориюны таңдаңыз. Содан кейін Export Files-ті басыңыз. Осыдан кейін collect.cfg, metadata.xml, sub.txt, org.txt, keywords.txt және AzList.txt бес файлы сол директорияға орналастырылады. Бұл файлдар жасау процесінде қолданылады, сондықтан оларды сәйкес директорияларға орналастыру өте маңызды.

Metadata.xml файлын import directory (импорт директория) атты директорияға, ал қалған файлдарды etc директориясына орналастырыңыз.

15 минутта 10 қадам жасаудан бастайық

Органайзермен жұмыс істеуді үйренудің ең жақсы тәсілі, жәй ғана кішірек кітапхана жасап көру. Егер сіздің 15 минут уақытыңыз болса, онда келесі 10 қадам сізге программамен толығырақ танысуға мүмкіндік береді.

Бастамас бұрын сізге:

- a. Құрамына DLS құрылымды Демо коллекция, өзінің бастапқы файлдарымен бірге кіретін Greenstone-ды орнату. Ескеріңіз, егер сіз өзіңіздің коллекцияңызға DLS коллекциясынан барлық 140 құжатты (Демо-коллекцияға енгізілген 11 дің орнына) қосқыңыз келсе, сіз кітапхананың осы DLS коллекциялары C:\Program\Files\gsdl\collect\demo және C:\Program\Files\gsdl\collect\dls директорияларына сәйкесінше орнатылады. Егер сіз Greenstone-дағы DLS-ті сіз орнатқан болсаңыз, енді оны тағы орнатқыңыз келсе онда сізге программаны алып тастап оны қайтадан өзіңізге керектілерін DLS-те көрсетіп орнатуыңыз қажет.
- b. Сіздің жаңа коллекцияңыз үшін (төменде біз оны “newcol” деп атаймыз) жіберуші мәзірдің *орындау* терезесінде келесі команданы енгізіп:
c:\program files\gsdl\bin\windows\build” newcol
- c. Алдыңғы қадамда жасалған бастапқы collect.cfg файлын, Демо-коллекция қолданысындағыға ауыстыру керек. Оны c:\program files\gsdl\collect\demo\etc\collect.cfg ны c:\program files\gsdl\collect\newcol\collect.cfg ке көшіру арқылы орындауға болады. Бұл қажет, себебі Демо-коллекция (және DLS-ке ұқсас барлық коллекциялар), бастапқы коллекция қолданбайтын арнайы опцияларды қолданады (жасаушыға арналған басшылықты қара).

Біз сізден осы нұсқауды қағазға шығарып алуды сұраймыз:

1. Органиайзерді ашып, dls мәліметтер базасын таңдаңыз және екі терезеде де “admin” паролін енгізіңіз (Вертикальді панельде collections кнопкасы белсенділенген болу керек, егер олай болмаса, онда оны қосыңыз).
2. Жаңа коллекция жасау үшін басты терезенің үстінде орналасқан көлденең панельден *New/Collection/Empty* командасын таңдаңыз. Оның атымен жасалым номерін беріңіз, мысалы “My First Collection” және жасалым номері “1.0”.
3. Кейбір мәтіндік атрибуттар үшін сіздің алдымен list of the possible values (мүмкін болатын мәндер тізімі) жасауыңыз керек.

Сіз құжат жазылған тіл мен оған қатысты ұйым атын келесі *New/Add-Modify languages* және *New/Organization* командаларын қолданып көрсете аласыз.

4. Жасаған коллекцияңыздың атына екі рет басыңыз.
5. Жоғарыдағы *Subjects* опциясына басыңыз (Ол Органайзерге кірген кезде үнсіздік бойынша белсенділендірілген); Содан кейін *Add subject*-ті басып *Add new subject* командасын таңдаңыз; енді *Subject title* (Рубрика атауы) терезесінде бірнеше рубрикаларды немесе тақырыпты әр рубриканы тергеннен кейін Enter-ді басып енгізіңіз. Жасаған рубрикалар санын көру үшін *Subjects* сөзінің алдынан “+” белгісін басыңыз.
6. *Documents view*-ті ашу үшін *Documents*-ке басыңыз. Бұдан кейін сіз коллекцияға құжаттарды келесі жолмен қоса аласыз.
 - а. Демо-коллекциядан (немесе DLS-тен егер бұл коллекция орнатылған болса) өзіңіздің жаңа коллекцияңызға қосуыңыз үшін, *Add documents from global list*-ті таңдаңыз. Өзіңізге қажет құжатты анықтап алыңыз (жоғарыда сипатталған “филтрация” функциясын қолданып) да оны коллекцияға қосыңыз. Құжатты қосқаннан кейін бұл құжаттың бастапқы файлдарын Демо-коллекцияның директориінен (с:\program files\gsdl\gsdl\collect\demo\import) тауып оны өзіңіздің жаңа коллекцияңыздағы import директориіне көшіріңіз. Мысалы, “Butterfly Farming in Papua New Guinea” құжатын өзіңіздің жаңа коллекцияңызға қосу үшін сіз өзіңіздің құжатыңыздың жұмыс номерін (job number) оны Органайзерге қосқаннан кейін анықтауыңыз керек. Бұл құжаттың жұмыс номері “b22bue”, сондықтан сіз “b22ue” директориін с:\program files\gsdl\collect\dls\import\ac01ne-ден с:\program files\gsdl\collect\newcol\import\ac01ne-не көшіруіңіз керек.
 - в. Жаңа құжаттарды (Демо-коллекциядағы құжаттардан басқа) өз коллекцияңызға қосу үшін *Add documents*-ті басып *Add new documents*-ті таңдаңыз. Құжат атауы мен жұмыс номерін (өз қалауыңызша), беттер санын, шығарушы ұйым, тіл және басқа да мәліметтерді енгізіңіз. Содан кейін сіз с:\program files\gsdl\collect\newcol\import директориясының ішінен осындай жұмыс номерімен жаңа ішкі директория жасауыңыз керек. Дәл осы директорияға сіз бұлақ файлдар мен импортталған құжаттың барлық графикалық файлдарын (бұл файлдар HTML-де, немесе Greenstone мен сәйкес келетін басқа форматтарда болуы керек (қолданушыға арналған басшылықты қара)) орналстыруыңыз керек.

7. *Subjects tab*-қа қайта оралсаңыз, өзіңіздің файлдарыңыздың төменгі терезеде бейнеленгенін көресіз. Құжатты таңдаңыз. Содан кейін осы құжатты классификациялаймын деген рубриkanı таңдаңыз. Содан кейін жоғары қаратылған көрсеткіші және қызыл кітап бар иконканы басыңыз. Құжатты классификациялағаннан кейін, сіз әліде оны бір рубрикадан басқасына тышқанның сол жақ кнопкасын пайдаланып тасымалдай аласыз. Рубрика бойынша ең аздегенде 6-32 құжаттарды классификациялауға тырысып көріңіз.
8. Алдыңғы қадамдарды құжаттар мен рубрикаларды көбірек қосып қайталаңыз. Сіз бұл үрдісті аяқтағаннан кейін барлық рубрикалар тізімінің аяқталғандығын, құжаттардың барлығының сәйкес рубрикалармен классификацияланғанына көз жеткізу үшін қарап шығыңыз.
9. Collection properties терезесін жауып, Вертикаль панельдегі Export Files кнопкасын басыңыз. Бұл басу Export Settings терезесінің пайда болуынаалып келеді. Display collections list-ке басып өз коллекцияңызды таңдаңыз, содан кейін Browse for folder-ге басып метамәліметтер файлын экспорттайтын директорияны таңдаңыз. Одан кейінгі құру процесстерінің метамәліметтер файлдарын экспорттау үшін Export files ті басыңыз.
10. Экспортталатын файлдарды өзіңіздің коллекцияңыздың сәйкес директорияларына көшіріңіз.
 - a. Экспортталатын collect.cfg файлында classify сөздерінен басталатын бірнеше жол бар. Сіз ең алдымен classify сөзінен басталатын жолдарды өзіңіздің жаңа коллекцияңыздан алып тастап, коллекцияңыздың конфигурациялық файлына (C:\program files\gsdl\collect.cfg) кесіп алып қыстыруыңыз керек.
 - b. Экспортталатын metadata.xml файлы C:\program\gsdl\collect\newcol\import directory ға көшірілуі қажет.
 - c. Экспортталатын AZList.txt, Keyword.txt, sub.txt, org.txt файлдары C:\program files\gsdl\collect\newcol\etc директориіне көшірілуі тиіс.

Енді өзіңіздің newcol коллекцияңыз тұрғызу үрдісіне дайын. Оны командалық жолдың көмегімен import.pl, build/pl командаларын қолданып тұрғызыңыз (жасаушыға арналған басшылықты қараңыз).

5.3 Құжаттар файлына тэгтер жасау

Егер сіз құрамын деген коллекцияның бастапқы құжаттары бөлімдер мен бөлімшелерге ұйымдастырылған болса, онда сіз бұны бастапқы (source files) файлында, Greenstone коллекциясы мазмұнын жасағанда, әрбір құжатқа сілтеме (active links) жасай алатындай етіп көрсетуіңіз керек. Бұны тақырып және бөлім тақырыпшасы метамәліметтерін белгілеу жолымен жасауға болады.

Бастапқы құжаттар дегеніміз әдетте COT үрдісінен кейін алынған Word-тың мәтіндік файлдар тобы, суреттермен қоса. Егер олар Word файлдары болса, онда олар Greenstone-ге Word plugin көмегімен қосыла алады. Альтернатива ретінде, олар сондай-ақ HTML форматына HTML plugin-ді қолданып түрлендіре алады.

Кез келген жағдайда, тақырыптар мен тақырыпшалар туралы метамәліметтер тэгтер арқылы келесі түрде көрсетілуі керек:

```
<!--
<Section>
<Description>
<Metadata name="Title"> Realizing human rights for poor
people: Strategies for achieving the international
development targets </Metadata>
</Description>
-->
(мұнда бөлім ішіндегілер жазылады)
<!--
</Section>
-->
```

<!--...--> маркерлері HTML дегідей комментарилерді көрсету үшін қолданады; осылайша, бұл бөлімдер тәгі құжаттарды форматтауға әсер етпейді. *Description* аумағында(<Description>, </Description> тэгтері арасында метамәліметтердің басқа түрлерін, көрсетуге болады, бірақ біз сипаттап отырған коллекциялар түрлері үшін олар қолданылмайды). Жоғарыда сипатталған жағдайлардың жалғасы ретінде бөлімдердің бөлімшелері болуы мүмкін:

```
( мұнда бөлім ішіндегілер жазылады)
<!--
<Section>
<Description><Metadata name="Title"> The international development
targets </Metadata>
</Description>
-->
(мұнда бөлім ішіндегілер жазылады)
<!--
</Section>
-->
(мұнда бөліс ішіндегілерден қалған бөлігі жазылады)
```

40 ЭЛЕКТРОНДЫ КОЛЛЕКЦИЯ ЖАСАУ

Мұндай функционалдыққа материалды HTML формасына түрлендіретін Word plugin көмегімен қолжеткізіледі, сондықтан да ол дәл солай Word (және RTF) тегі метамәліметтерді анықтауға қолданыла алады (Word тағы құжаттарды HTML-ге түрлендіргенде, әдетте HTML-дің арнайы“<“and”>“ белгілерін нейтрализациялауға көп уақыт бөлінеді). Байқасаңыз Word файлдарында да “<!- -”and”- ->” тура жоғарыдағыдай формат қолданылады. Шрифтер мен тәттік белгілер ескерілмейді.