

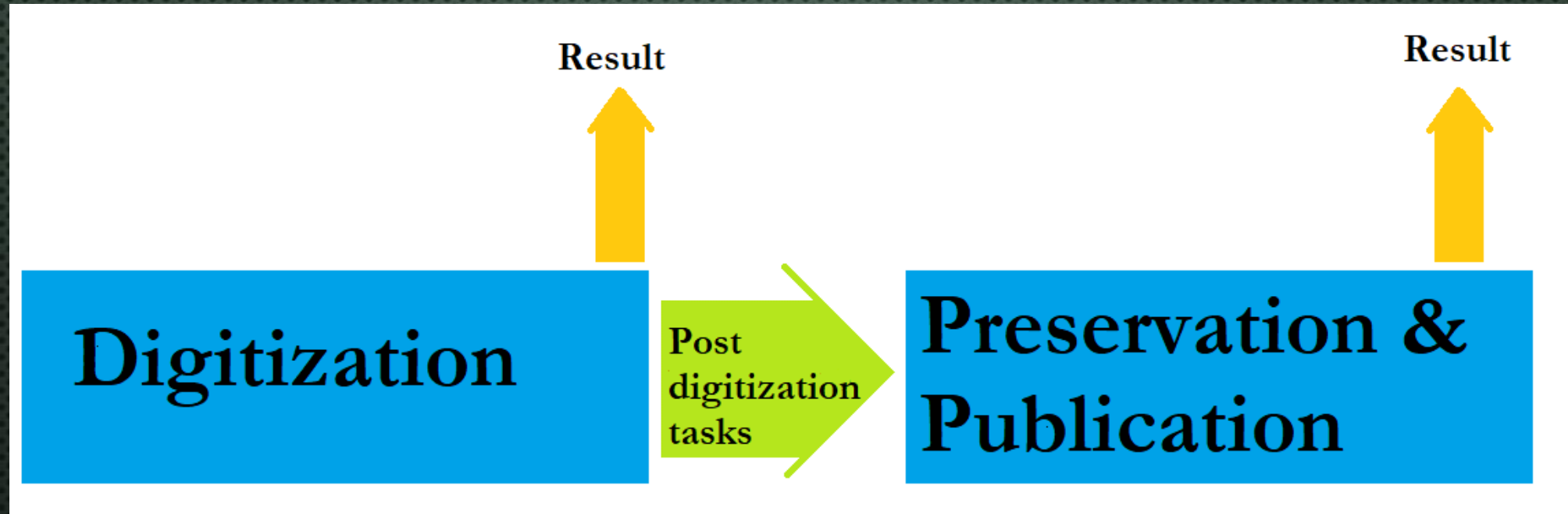


AUTOMATED POST-DIGITIZATION SYSTEM

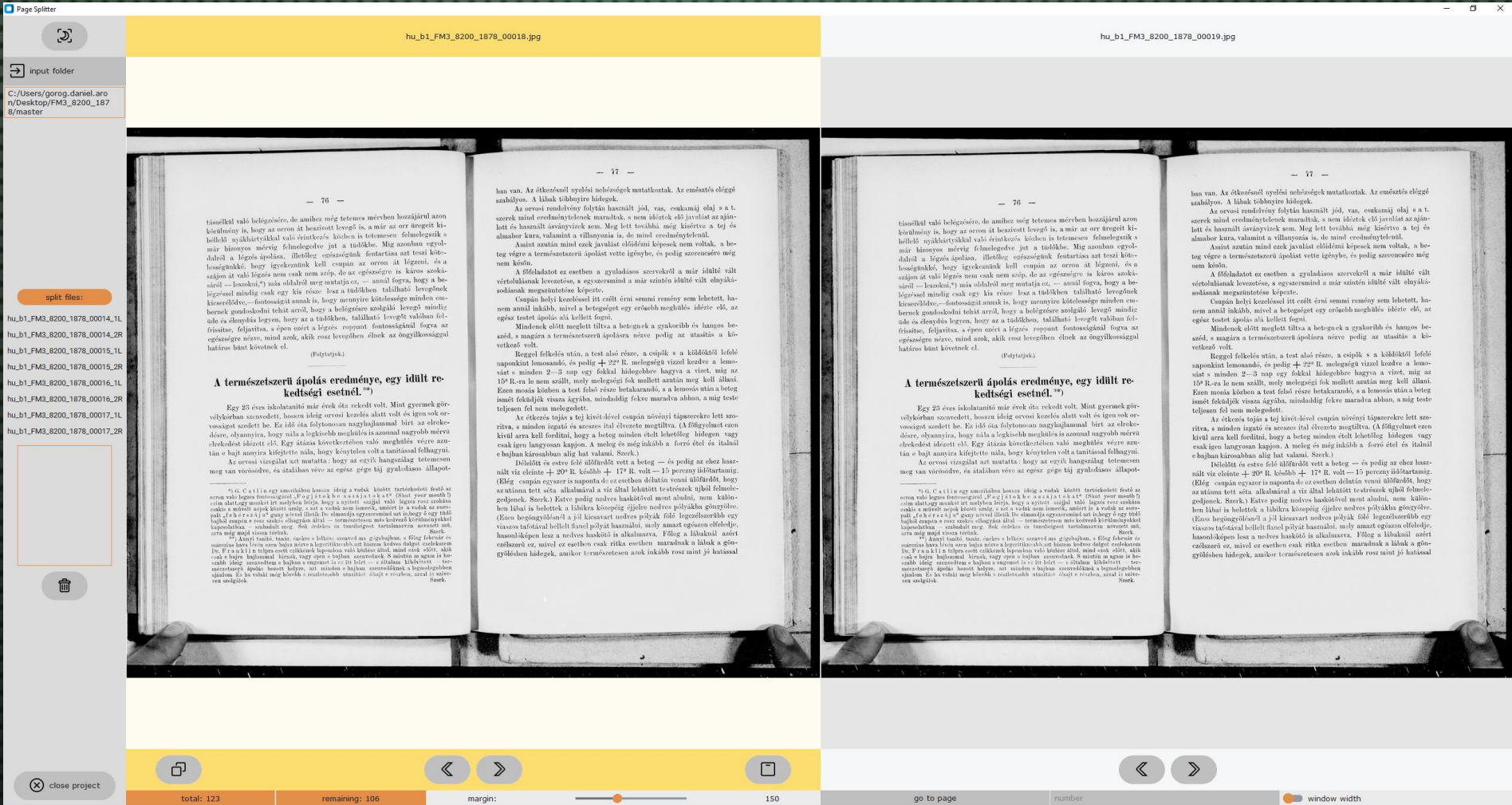
DÁNIEL GÖRÖG, GÁBOR KLINGER, MÁTYÁS RÉNYI
MNMKK OSZK
DIGITAL PRESERVATION DEPARTMENT

■ WHAT IS POST-DIGITIZATION?

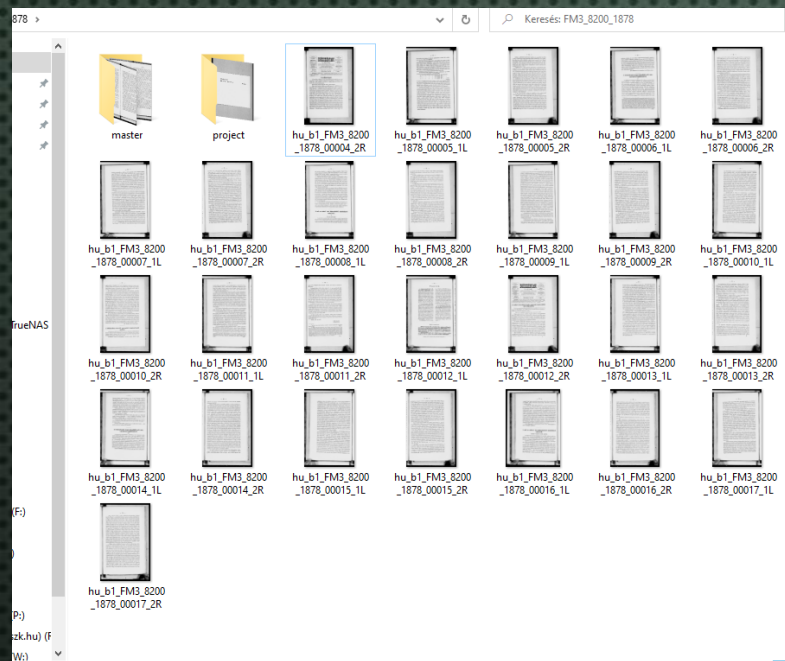
- NORMALIZATION OF FOLDER STRUCTURE AND FILE NAME
- CORRECTION OF CERTAIN TYPES OF FAULTS
- IN CASE OF PERIODICALS SCANNED FROM MICROFILMS: SPLITTING DOUBLE PAGES TO SINGLE PAGES, AND SORTING THEM AS ISSUES
- LABOUR INTENSE TASKS WITH LITTLE ADDED VALUE
- STUCK BETWEEN DIGITIZATION AND PRESERVATION



PAGESPLITTER: THE SOLUTION

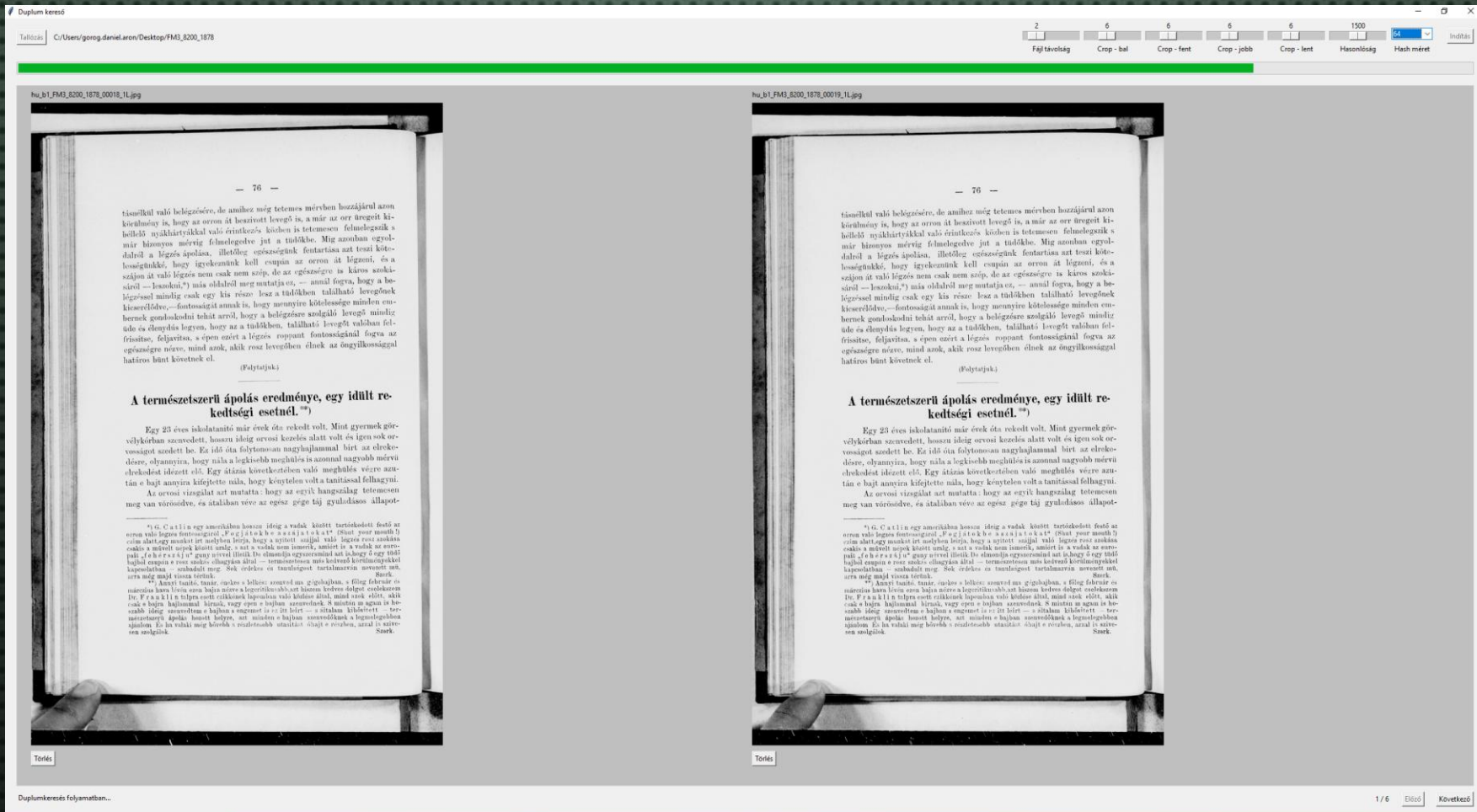


PAGESPLITTER: OUTPUT



- SIMPLICITY
- EASY HANDLING
- CUSTOMIZED FOR US

DUPLUMFINDER: THE METHOD

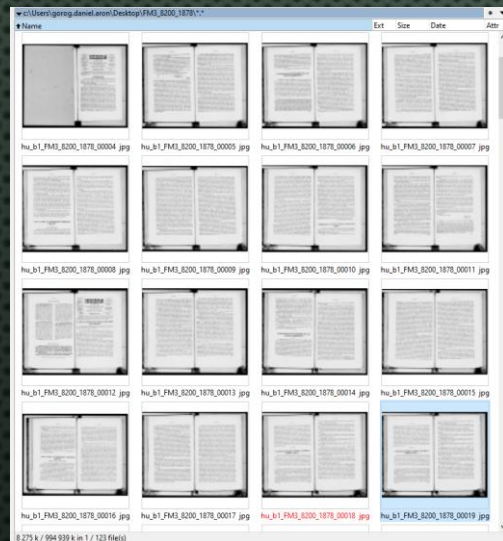


DUPLUMFINDER: THE METHOD

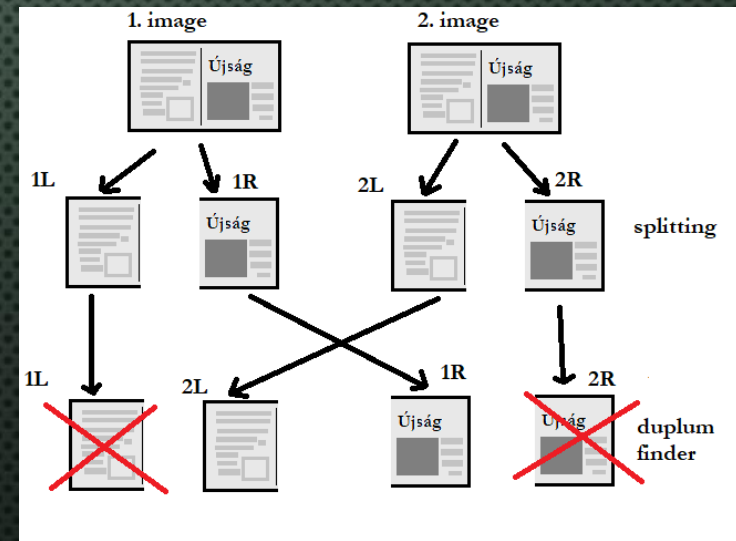
- WORK DONE BY THE COMPUTER, ERGONOMICS
- CUSTOM HASH NUMBERS, SENSITIVITY AND MARGINS
- FINDS TRIPLETS, QUADRUPLETS TOO
- OPTION NOT TO DELETE
- SAFETY: NOT THE COMPUTER DELETES, BUT THE OPERATOR

DUPLUMFINDER: PROBLEMS WITH THE PROCESS

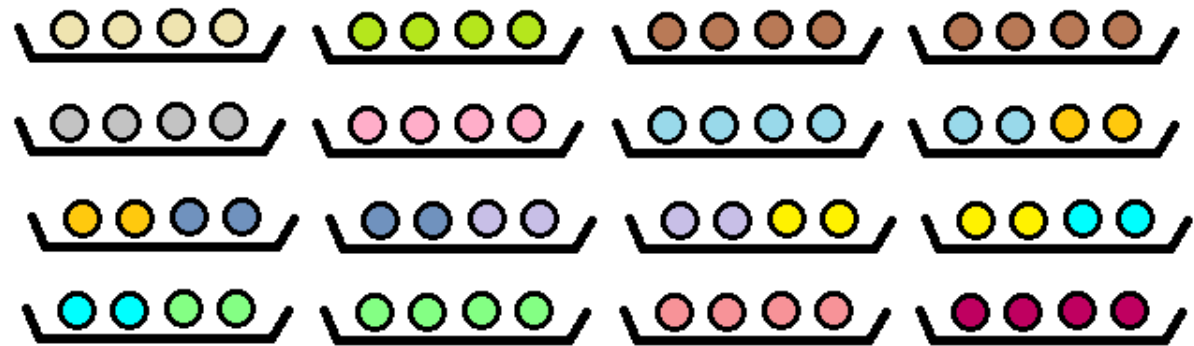
ERGONOMIC ISSUES



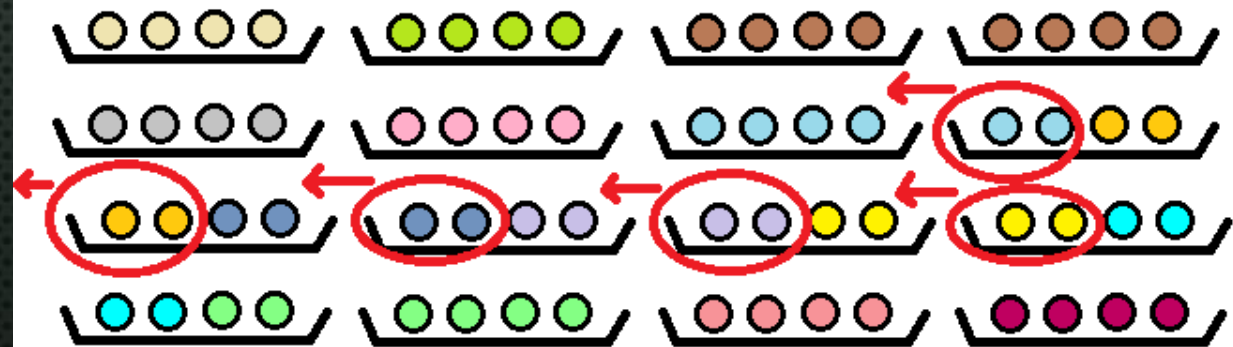
RELATIVE ORDER



SEGMENTATION: FIRST TRY



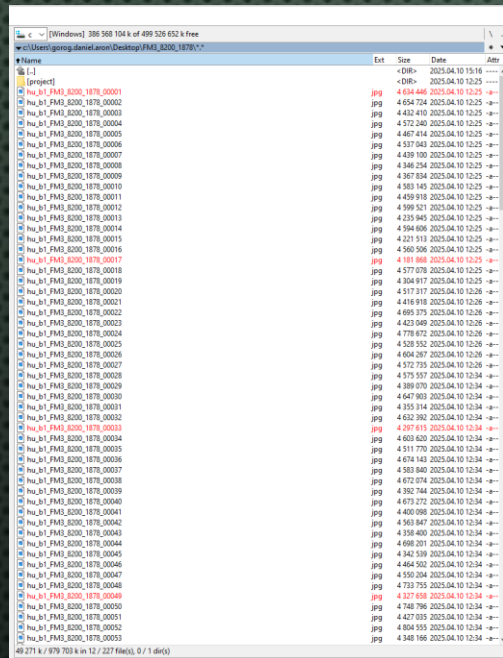
SEGMENTATION: FIRST TRY



SEGMENTATION: ARTIFICIAL INTELLIGENCE

- THE ARTIFICIAL INTELLIGENCE COULD RECOGNIZE THE COVER PAGES (BY EXAMPLE GIVEN)
- THE FOLLOWING PAGES WOULD BE PUT UNDER THE COVER PAGES, THUS THE ISSUE RECREATED
- STRONG HUMAN CONTROL
- AI IS NOT CONSISTENT IN RECOGNIZING THE COVER PAGES

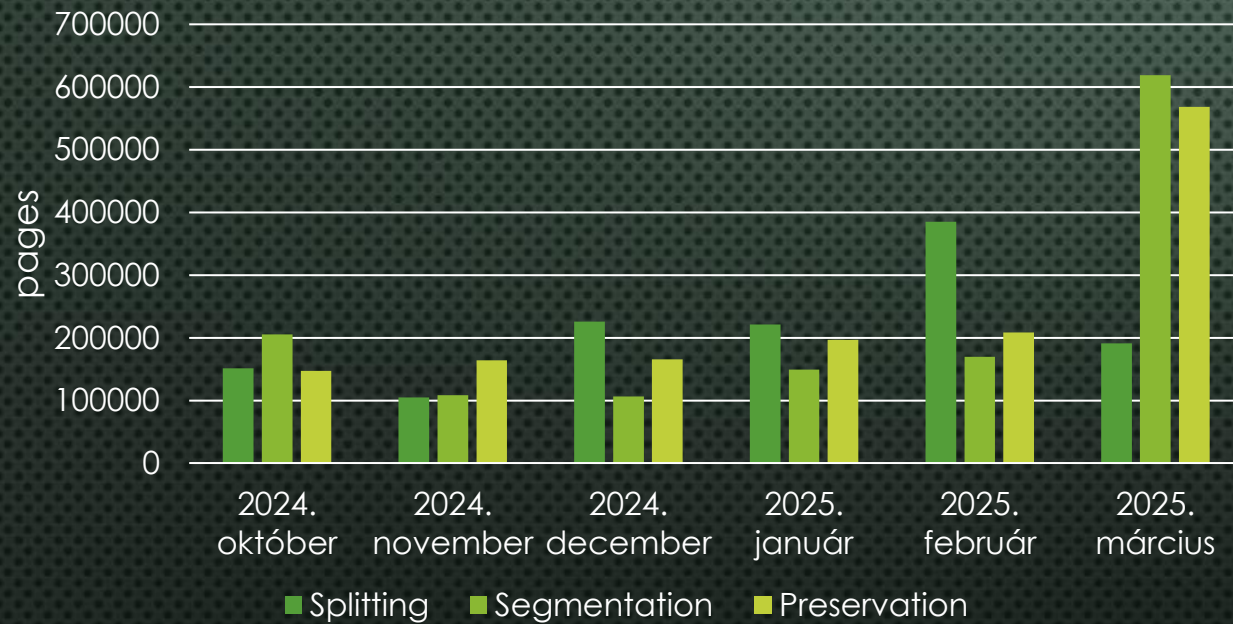
SEGMENTATION: MMM SEMI-AUTOMATIC METHOD



Name	Ext	Size	Date	Attr
[...]				
img_b1_FM3_8200_1878_00001	.jpg	4 634 446	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00002	.jpg	4 654 724	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00003	.jpg	4 432 410	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00004	.jpg	4 572 240	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00005	.jpg	4 467 414	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00006	.jpg	4 537 043	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00007	.jpg	4 439 100	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00008	.jpg	4 340 254	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00009	.jpg	4 367 834	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00010	.jpg	4 583 145	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00011	.jpg	4 459 918	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00012	.jpg	4 599 521	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00013	.jpg	4 235 945	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00014	.jpg	4 594 689	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00015	.jpg	4 221 513	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00016	.jpg	4 560 506	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00017	.jpg	4 181 848	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00018	.jpg	4 577 078	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00019	.jpg	4 304 917	2025.04.10 12:25	-r--
img_b1_FM3_8200_1878_00020	.jpg	4 517 517	2025.04.10 12:26	-r--
img_b1_FM3_8200_1878_00021	.jpg	4 416 918	2025.04.10 12:26	-r--
img_b1_FM3_8200_1878_00022	.jpg	4 695 375	2025.04.10 12:26	-r--
img_b1_FM3_8200_1878_00023	.jpg	4 423 549	2025.04.10 12:26	-r--
img_b1_FM3_8200_1878_00024	.jpg	4 778 672	2025.04.10 12:26	-r--
img_b1_FM3_8200_1878_00025	.jpg	4 538 552	2025.04.10 12:26	-r--
img_b1_FM3_8200_1878_00026	.jpg	4 604 247	2025.04.10 12:26	-r--
img_b1_FM3_8200_1878_00027	.jpg	4 572 755	2025.04.10 12:26	-r--
img_b1_FM3_8200_1878_00028	.jpg	4 575 557	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00029	.jpg	4 389 070	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00030	.jpg	4 647 903	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00031	.jpg	4 555 314	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00032	.jpg	4 632 582	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00033	.jpg	4 297 615	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00034	.jpg	4 460 620	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00035	.jpg	4 511 770	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00036	.jpg	4 674 143	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00037	.jpg	4 583 840	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00038	.jpg	4 672 074	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00039	.jpg	4 392 744	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00040	.jpg	4 672 272	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00041	.jpg	4 400 098	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00042	.jpg	4 563 947	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00043	.jpg	4 358 400	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00044	.jpg	4 698 201	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00045	.jpg	4 342 539	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00046	.jpg	4 464 582	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00047	.jpg	4 550 204	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00048	.jpg	4 733 755	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00049	.jpg	4 577 618	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00050	.jpg	4 748 796	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00051	.jpg	4 427 055	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00052	.jpg	4 804 915	2025.04.10 12:34	-r--
img_b1_FM3_8200_1878_00053	.jpg	4 348 166	2025.04.10 12:34	-r--

- COVER PAGES ARE SELECTED BY THE HUMAN OPERATOR
- COMPUTER SCRIPTS ORGANIZE THE FOLLOWING PAGES UNDER THE COVER
- NO ROOM FOR AI ERROR
- APPROX 8X THE SPEED

SEGMENTATIO: RESULTS



- NEW SOLUTIONS INVENTED
- FAST, SAFE AND ERGONOMIC METHODS
- TENDENCY TO BE FASTER IN OUR PRESERVATION WORK

PRINCIPLES AND CONSEQUENCES

- COMPUTER BASED WORK UNDER HUMAN CONTROL
- COMPUTER DOES THE BIG-SIZED, MONOTONOUS TASKS
- THE HUMAN OPERATOR DOES THE CREATIVE OR RESPONSIBLE TASKS
- SUCCESSFUL AUTOMATIZATION, BIG NUMBERS

THANK YOU FOR YOUR
ATTENTION!

DÁNIEL GÖRÖG, GÁBOR KLINGER, MÁTYÁS RÉNYI
MNMKK OSZK
DIGITAL PRESERVATION DEPARTMENT