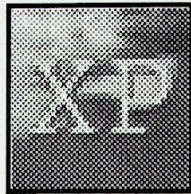


# Xtrapress

By  
Subhan Ali Khan



Supervised  
By

Mr. Zarrar Javaid

A report is submitted to the department of Computer Science,  
Bahria Institute of Management and Computer Science, Islamabad

In partial fulfillment of requirement for the degree of MCS

---

**Department of Computer Sciences**  
Bahria Institute of Management and Computer Science, Islamabad  
University of Peshawar, Peshawar



## Dedication:gements

This Project is dedicated to my wife. Mr. Lee Yac, Professor of Computing Systems at MSU, for sparing time and patience to teach the basics of Block Sorting and File Archiving. Mr. Zamar, the project supervisor, for invaluable aid and advice. Mr. Ehsan Khan, the wizard of Delphi, for the debugging of the project. All the subject tutors who were all so understanding.

All who have responded to the beta testing program and submitted feedback make especially no particular order. Jamar Khalid, Raheem, for their support, for their encouragement and understanding for those moments when God was always there when I needed him.



# Acknowledgements

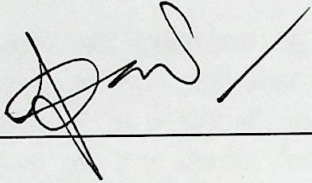
I would like to extend my thank to Mr. Lee Kao, Professor of Computing Dynamics at MSU, for sparing time and patience to teach the basics of Block Sorting and File Archiving. Mr. Zarrar, the project supervisor, for invaluable aid and advice. Mr Ehsan Khan, the wizard of Delphi, for the debugging of this project. All the subject tutors who were all so understanding.

All who have responded to the beta testing program and submitted benchmark results, especially (in no particular order), Jamal Khalid, Raihan Ahmed. My parents for their encouragement and understanding for those sleepless nights. God, who was always there when I needed him.

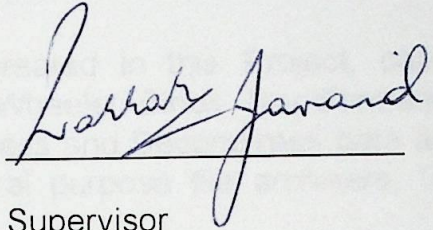


# Certificate

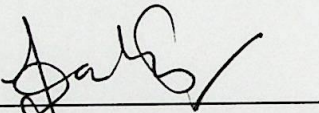
We accept the work contained in this report as a confirming to the required standard for the partial fulfillment of the degree of BCS/BSE/BCE/BBA/MCS/MBA in the subject of \_\_\_\_\_



Head of Department

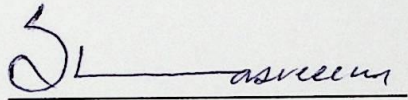


Supervisor



Internal Examiner

Fazal Wahid



External Examiner

(Prof. Dr. Tasneem Shah)



# Abstract

There is a need for better and more efficient Data Compression Applications. There are quite a few available commercially, but most of them are based on Dictionary Based Compression Techniques.

This project tries to differentiate from the conventional way of compressing data, and uses a Compression Technique which is far more superior, efficient and closer to the state of the art Statistical Compressor for Super Computers.

For such an exercise, a File Archiver is created in this Project, called Xtrapress, which will use the Burrows and Wheeler Block Transformation Algorithm. This Archiver will be able to Compress and Decompress data and will have all the basic features of the general purpose file archivers, like WinZip or WinRAR.

2.17. DISCUSSION	24
2.18. THE RESULTS OF WINRAR	249
2.19. MOTIVATION	251
2.20. USING THE OPPORTUNITY	252
2.2. CONCLUSION	253
2.3. THE BENEFITS OF BWT	253
<b>3. PROPOSED SYSTEM</b>	<b>3-15</b>
3.1. WHAT IS XTRAPRESS	3-15
3.2. WHY USE XTRAPRESS	3-15
3.3. THE NAME XTRAPRESS	3-15
3.4. THE OPERATIONS	3-17
3.5. FEATURES	3-17
3.6. REQUIREMENTS	3-17
3.6.1. DEVELOPMENT	3-18
3.6.2. SYSTEM REQUIREMENTS	3-18
3.6.3. RECOMMENDED ENVIRONMENT	3-18
<b>4. IMPLEMENTATION</b>	<b>4-19</b>
4.1. DEVELOPMENT STRATEGY	4-20
4.2. DESIGN STRUCTURE	4-21
4.3. CLASSES	4-25
4.3.1. MODEL AND CLASSES	4-25
4.3.2. CLASS LIBS AND THEIR FULL NAMES	4-26
4.3.3. HOW TO USE THE CLASSES	4-26
4.3.4. CONSTRUCTORS	4-26
4.3.5. CLASS METHODS	4-27
4.3.6. DATA CLASS	4-28
4.4. ALGORITHM	4-31
4.4.1. THE BURROWS AND WHEELER BLOCK TRANSFORMATION	4-31
4.4.2. THE BURROWS AND WHEELER BLOCK TRANSFORMATION	4-37
4.4.2.1. Description of Development Techniques	4-37
4.4.2.2. Name	4-37



# Table of Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>1-1</b>
1.1	ARCHIVING AND DATA COMPRESSION	1-2
1.2	AIM AND OBJECTIVE	1-2
1.3	THE PROBLEM	1-3
1.4	THE SOLUTION	1-3
1.5	BWT ALGORITHM	1-3
1.6	SCOPE	1-5
<b>2</b>	<b>LITERATURE SURVEY</b>	<b>2-6</b>
2.1	THE BURROWS-WHEELER TRANSFORM	2-7
2.1.1	BWT BASICS	2-7
2.1.2	OBSERVATIONS	2-9
2.1.3	THE ROSETTA VECTOR	2-10
2.1.4	MOVING ON	2-11
2.1.5	USING THE OPPORTUNITY	2-12
2.2	CONCLUSION	2-13
2.3	THE BENEFITS OF BWT	2-13
<b>3</b>	<b>PROPOSED SYSTEM</b>	<b>3-15</b>
3.1	WHAT IS XTRAPRESS	3-16
3.2	WHY USE XTRAPRESS	3-16
3.3	THE NAME XTRAPRESS	3-16
3.4	XTRAPRESS ICON	3-17
3.5	FEATURES	3-17
3.6	REQUIREMENTS	3-17
3.6.1	DEVELOPMENT	3-18
3.6.2	MINIMUM REQUIREMENTS:	3-18
3.6.3	RECOMMENDED REQUIREMENTS:	3-18
<b>4</b>	<b>IMPLEMENTATION</b>	<b>4-19</b>
4.1	DEVELOPMENT STRATEGY	4-20
4.2	DESIGN STRUCTURE	4-21
4.3	CLASSES	4-22
4.3.1	MODULE AND CLASSES	4-22
4.3.2	CLASS LISTING AND THEIR FULL NAMES	4-24
4.3.3	HOW TO USE THE CLASSES	4-26
4.3.4	CLASS USAGE	4-26
4.3.5	CLASS HIERARCHY	4-28
4.3.6	DATA FLOW	4-30
4.4	ALGORITHMS	4-31
4.4.1	THE RUN LENGTH ENCODER AND DECODER	4-31
4.4.2	THE BURROWS WHEELER FORWARD TRANSFORMATION	4-32
4.4.2.1	Description of the Forward Transformation	4-32
4.4.2.2	Sorting	4-33



4.4.2.3	The SwapBlocks management system	4-33
4.4.3	THE BURROWS WHEELER REVERSE TRANSFORMATION	4-34
4.4.3.1	Description of the Reverse Transformation	4-34
4.4.4	THE MOVE TO FRONT ENCODER AND DECODER	4-35
4.4.5	THE ARITHMETIC ENCODER AND DECODER	4-35
4.4.5.1	Overview of arithmetic encoding process:	4-35
4.4.5.2	The Unstructured group of components:	4-36
4.4.5.3	TFileStrucAriEncoder	4-37
4.4.5.4	TFileStrucAriDecoder	4-37
4.4.5.5	TStrucAriEncoder	4-37
4.4.5.6	TStrucAriDecoder	4-38
4.4.5.7	THeadAriModel and TGroupAriModel	4-38
<b>4.5</b>	<b>ENGINE</b>	<b>4-40</b>
4.5.1	THE ARCHIVE MANAGER	4-40
4.5.1.1	Opening Archives	4-40
4.5.1.2	Adding Files	4-40
4.5.1.3	Extracting Files	4-41
4.5.1.4	Deleting Files	4-41
4.5.1.5	Modifying file Attributes	4-42
4.5.1.6	Other Considerations	4-42
4.5.1.7	To Use	4-42
4.5.2	THE ARCHIVE FILE CLASS	4-42
4.5.3	THE BIT FILE CLASS	4-43
4.5.4	THE SMART BUFFERED FILE STREAM CLASS	4-43
4.5.5	THE CENTRAL DIRECTORY CLASS	4-44
4.5.6	THE CRC 32 CLASS	4-44
4.5.6.1	To Use	4-44
4.5.7	ARCHIVE FILE STRUCTURE	4-44
4.5.7.1	Data Types and Definition	4-44
4.5.7.2	General Format of an Xtrapress Archive	4-45
4.5.7.3	Signatures	4-45
4.5.7.4	Headers	4-45
4.5.7.5	Overall Xtrapress Format	4-46
<b>4.6</b>	<b>ADDITIONAL NOTES</b>	<b>4-47</b>
4.6.1.1	Long file names	4-47
4.6.1.2	Sequential access storage	4-47
4.6.1.3	Command Line Processing	4-47

---

## **5 TESTING / EVALUATION** **5-48**

<b>5.1</b>	<b>TESTING STRATEGY</b>	<b>5-49</b>
<b>5.2</b>	<b>DIFFERENT ENVIRONMENTS</b>	<b>5-49</b>
<b>5.3</b>	<b>TEST DATA</b>	<b>5-49</b>
<b>5.4</b>	<b>COMMAND LINE TESTING</b>	<b>5-49</b>
<b>5.5</b>	<b>INPUT FILE TEST DATA</b>	<b>5-50</b>
<b>5.6</b>	<b>TESTING DURING DEVELOPMENT: (ALPHA TESTING)</b>	<b>5-51</b>
5.6.1	DELPHI STANDARD DEBUGGING TOOLS:	5-51
5.6.2	ASSERTIONS:	5-51
5.6.3	DEBUGGING STATEMENTS:	5-51
5.6.4	TESTING AIDS:	5-51
5.6.5	HIDDEN BUTTONS:	5-52
5.6.6	DEBUG FORM:	5-52
5.6.7	LOGICAL TESTING:	5-52
5.6.8	REVERSIBILITY TESTING:	5-52



5.6.9	SORT TESTING:	5-52
5.6.10	COMPARING AND VALIDATING BLOCKS:	5-53
5.6.11	MULTIPLE BLOCKS:	5-53
5.6.12	BOUNDS CHECKER:	5-53
5.6.13	BACKING UP:	5-53
5.6.14	FILE TYPES:	5-53
5.6.15	INTENSIVE TESTING OR 'MASSIVE' TESTING:	5-54
<b>5.7</b>	<b>TESTING THE FINAL PROGRAM: (BETA TESTING)</b>	<b>5-54</b>
<hr/>		
<b>6</b>	<b>CONCLUSION</b>	<b>6-55</b>
6.1	EVALUATION	6-56
6.2	CLOSING REMARKS	6-56
<hr/>		
<b>7</b>	<b>FUTURE DEVELOPMENT</b>	<b>7-57</b>
7.1	REDUCTION IN MEMORY USAGE	7-58
7.2	BETTER COMPRESSION	7-58
7.3	ADDING OF FOLDERS	7-58
7.4	DISK SPANNING FACILITY	7-58
7.5	DRAG AND DROP	7-58
7.6	ERROR RECOVERY	7-58
7.7	CONFIGURATION SAVING	7-59
7.8	MORE COLOURFUL INTERFACE	7-59
7.9	BETTER INTERFACE FOR ADDING AND EXTRACTING FILES	7-59
7.10	AUTOMATIC CONVERSION FROM OTHER ARCHIVE FORMATS	7-59
<hr/>		
<b>8</b>	<b>BIBLIOGRAPHY</b>	<b>8-60</b>
8.1	BLOCK SORTING	8-61
8.2	ARITHMETIC MODELLING AND CODING	8-61
<hr/>		
<b>9</b>	<b>APPENDICES</b>	<b>9-62</b>
9.1	APPENDIX A – INSTALL / UNINSTALL	9-63
9.1.1	INSTALLING XTRAPRESS	9-64
9.1.2	UNINSTALLING XTRAPRESS	9-64
9.2	APPENDIX B – USER INTERFACE	9-65
9.2.1	INTRODUCTION TO THE XTRAPRESS MAIN SCREEN	9-66
9.2.1.1	Add a file.	9-66
9.2.1.2	Extract this file.	9-66
9.2.1.3	Change the file name.	9-66
9.2.1.4	Delete this file.	9-67
9.2.2	INTERFACE DESIGN CONSIDERATIONS	9-68
9.2.3	DIALOG POSITIONS	9-70
9.2.4	MESSAGE BOX	9-70
9.2.5	OPENING AND CREATING ARCHIVES	9-72
9.2.6	SELECTING FILES	9-74
9.2.7	ADDING FILES	9-75
9.2.8	EXTRACTING FILES	9-76
9.2.9	DELETING FILES	9-77
9.2.10	MODIFYING FILE ATTRIBUTES	9-78



9.2.11	SORTING FILES	9-79
9.2.12	CONFIGURATION DIALOG	9-80
9.2.13	THE COMPRESSION STATISTICS DIALOG	9-81
9.2.14	THE PROGRAM STATISTICS DIALOG	9-82
9.2.15	ABOUT DIALOG	9-83
<b>9.3</b>	<b>APPENDIX C – TROUBLESHOOTING</b>	<b>9-84</b>
9.3.1	OPENING FILES	9-85
9.3.2	DURING OPERATIONS	9-85
<b>9.4</b>	<b>APPENDIX D – BUGS AND LIMITATIONS</b>	<b>9-86</b>
9.4.1	KNOWN BUGS OR QUIRKS	9-87
9.4.2	LIMITATIONS	9-87
<b>9.5</b>	<b>APPENDIX E – BENCHMARKS</b>	<b>9-88</b>
9.5.1	THE CALGARY CORPUS	9-89
9.5.2	TABLE 1 – STRUCTURED VS. UNSTRUCTURED MODEL	9-90
9.5.2.1	Structured model	9-90
9.5.2.2	Unstructured model	9-90
9.5.3	TABLE 2 – COMPRESSION OF THE CALGARY CORPUS	9-92
9.5.4	TABLE 3 – COMPRESSION OF INDIVIDUAL FILES	9-94
9.5.5	TABLE 4 – INCREASING THE BLOCK SIZE	9-96
9.5.6	ANALYSIS OF BENCHMARKS	9-97