

INDICE

Indice

RIASSUNTO.....	1
INDICE.....	2
CAPITOLO 1: INTRODUZIONE.....	4
Le fasi del progetto e le metodologie adottate.....	5
Il contenuto della tesi.....	9
CAPITOLO 2: IL CONTESTO.....	11
2.1 Analisi del mercato.....	11
2.2 Analisi del settore.....	15
2.3 Possibili scenari futuri.....	16
2.4 Le associazioni di categoria.....	18
2.4.1 Assoprovider.....	18
2.4.2 AIIP - Associazione Italiana Internet Provider.....	19
2.4.3 Il consorzio VoIPEX.....	20
2.5 Il caso di studio: l'azienda CWNET.....	21
2.5.1 Una impresa proattiva.....	21
2.5.2 Organizzazione aziendale.....	22
2.5.3 Cheapnet.it: il sito e la storia.....	25
2.6 Conclusioni.....	27
CAPITOLO 3: IL PROGETTO FORMATIVO.....	28
3.1 Il sistema informativo.....	28
3.2 Le fasi del progetto.....	29
3.3 Estrazione, pulizia, trasformazione e caricamento dei dati.....	30
3.3.1 Estrazione dei dati.....	30
3.3.2 Le tabelle del database.....	31
Trasformazione e caricamento dati.....	37
3.3.3 Il file degli accessi al sito.....	40
Trasformazione del file di log.....	41
CAPITOLO 4: IL DATA WAREHOUSE.....	43
4.1 Introduzione al data warehousing.....	43
4.2 L'approccio metodologico.....	44
4.2.1 Approccio top-down.....	45
4.2.2 Approccio bottom-up.....	46
4.2.3 Approccio incrementale.....	46
4.3 Metodi di accesso ai dati.....	47
4.4 I cubi OLAP.....	48
4.5 Il data warehouse di CWNet.....	50
Obiettivo 1: Analisi dell'efficacia delle campagne promozionali.....	51
Obiettivo 1: I risultati conseguiti.....	55
Obiettivo 2: Analisi delle vendite.....	58
L'analisi del portafoglio prodotti.....	59
La matrice BCG.....	59
Il portafoglio prodotti di CWNet.....	62
Il data mart.....	65
Obiettivo 2: I risultati conseguiti.....	67
4.6 Conclusioni.....	72

CAPITOLO 5: DATA MINING.....	74
5.1 Comprensione del problema e definizione degli obiettivi.....	74
5.2 Il metodo CRISP-DM.....	77
5.3 Dalla market basket analysis alla recommendation system.....	79
5.4 Data mining per il web.....	86
5.5 Analisi del traffico web.....	87
Preprocessing.....	87
WOS: Web Object Store.....	89
Ricostruzione delle sessioni.....	89
Risultati ottenuti: il grafo del traffico web.....	90
5.6 Analisi del comportamento degli utenti.....	99
WUMPrep: comprensione, pulizia e trasformazione dei dati.....	99
WUM: identificazione e ricostruzione delle sessioni.....	102
Risultati ottenuti.....	105
5.7 Conclusioni.....	111
CAPITOLO 6: CONCLUSIONI.....	112
APPENDICE.....	115
Esempio de file di accessi log per il sito Cheapnet.it.....	115
Tassonomia.....	116
Le macroaree del sito Cheapnet.it.....	117
Confronto tra MySQL e SQL Server 2000.....	118
Metodologia per risalire all'ISP di provenienza del contatto.....	120
Bibliografia.....	122
Sitografia.....	123
RINGRAZIAMENTI.....	124