

eBook Authoring System concept design and development

Mohamed Menacer, Amar Arbaoui

NOOR Research Centre, College Computer Science & Engineering,
Taibah University, Madinah Saudi Arabia

Abstract— This paper presents the design and development of an authoring system for creating and building an application program for an eBook package, or an eBook portfolio, that would include metadata, meta-content, and other related materials and resources that are stored in a dedicated database. This system is basically a digitization process of the different phases of eBook authoring and development, as well as a structure collection of various types of content and information in a form of hypertext documents, multimedia, and presentations. Although the open eBook format standard has been developed in the late 1990s, it has been mainly driven by commercial and marketing issues in mind rather than academic or/and technical focus. The standard is to some extent limited, in the opinion of the authors, in some technical aspects, i.e. advanced search and indexations, as well as in the subject matter areas such as education where content requires more and diverse representations of the academic content. The authoring system, being proposed, has been designed with academic books in mind, it is still very much suitable for any type of documentation or eContent such as articles, papers, reports, and presentations. This approach of eBook and eContent authoring and development is very useful for online search, manipulation, indexing, distribution and even marketing purposes. Based on the proposed content structured content, web content can be easily and reliably accessible for search and information retrieval. Furthermore, Natural Language Processing techniques such as text mining, indexing, tagging, information retrieval for online content document processing would be implemented more reliable and efficiently. Overall design concepts and development of the authoring system approach is presented in this paper with the different phases involved.

Keywords: Authoring system, eBook, eContent, Meta-content, Digital content creation, Multimedia, eContent distribution, eLearning.

Introduction

Digital reading provides new opportunities for readers, publishers, and authors to discover, consume, and connect in different electronic book formats. An electronic book, or eBook, is a digital version of a printed book that contains different type of contents such as text, figures, schematics, tables, and images readable on computers-like devices, eReaders, Tablets, and Smartphone [1]. According to PricewaterhouseCoopers, one of the world's largest providers of business consulting services, and based on surveys of experts in the field, the book market is in a transition phase where readers are making the leap from paper books to eBooks and eReaders, which is poised to make significant changes and revolutionize the book industry forever [2]. Although there are a number of eBook software and tools available in the market [3], there are a limited number of eBook authoring systems that would allow users to create eBooks and related multimedia application capable of manipulating multimedia objects and resources [4]. Furthermore, not all eBooks are well structured for accurate search and efficient delivery and marketing through the web or/and to be used by third party applications or educational information systems such as eLearning and Virtual Learning Environment. By providing better digital book structure and more content details, and especially meta-content, would allow for better search and indexing as well as further improvement and integration with third party authoring applications and systems [5,6].

The aim of this paper is to present a systematic approach for authoring eBooks by collecting and digitizing the most relevant content and information in a structured manner of academic as well as general purpose books, or similar documents, for online and offline efficient usage, delivery and dissemination. The above work is being implemented under a research project, which is at its initial stages.

Digital book concepts

Due to the lack of a standard for packaging and selling eBooks, a consortium was formed in the late 1990s that developed the Open eBook standard (OEB), currently Open Publication Structure (EPUB), of a single source-document that can be handled by many book-reading software and hardware platforms [7]. The standard defined subsets of XHTML and CSS as a set of multimedia formats, and an XML schema for a "manifest" in order to list the components of a given eBook that identify a table of contents, front cover, and so on [8].

EPUB is an open specification for digital books, emerging as the de facto standard, based on familiar technologies like XML, CSS, and XHTML, and EPUB files can be read on portable devices, mobile phones, and desktop computers [9]. Figure 1 [10], is a generic representation of an eBook user visual interface which consists of Control Bar, Search Bar, Main Panel and navigation or Left Panel. Each eBook contains usually virtual file directories for each panel. The main panel displaying the main content of the eBook, and the navigation panel providing its structure as well as access to the different metadata and other relevant info [10].

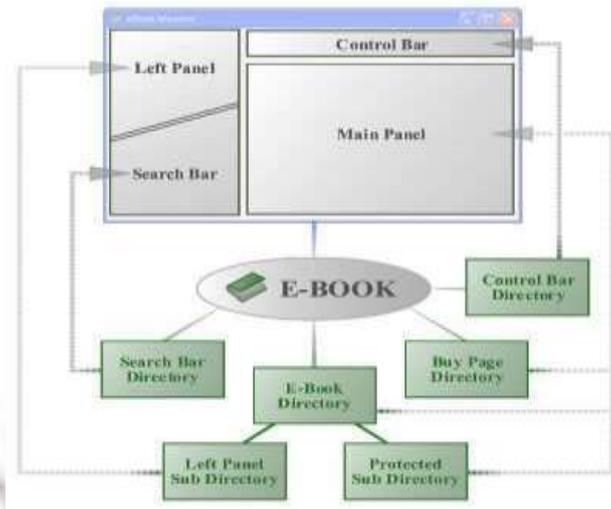


Figure 1. eBook Structure concept [10].

Lately, many eBook publishers began distributing books in different file and proprietary file formats, due to the very attractive and promising commercial future prospects that electronic books offer [11,12,13]. Digital book technical development has been mainly driven by commercial/marketing issues in mind rather than academic or/ and technical focus. The focus was on the output or delivery part of the eBook, which is still somewhat limited, in the opinion of the authors, in some technical aspects, i.e. advanced search and indexing, as well as in the subject matter areas such as education where content requires more and diverse representations of the academic content [14]. Thus the main idea of developing an eBook authoring system which is rich and diverse in content type as well as widely and efficiently disseminated. The structure and format, of the proposed eBook authoring system, would be very adequate for all subject matters especially for academic books and documentations.

Proposed eBook Content Structure

Figure 2 illustrates the various content and information that would constitute an eBook content structure. Five major categories of meta details has been considered:

- eBook Metadata: Title, Author(s) details and biography, Publishing details, book description, book abstract, keywords.
- eBook Meta-content: Chapter, Paragraph, sub-paragraph, table of content, figures, diagrams, tables, Appendix.
- eBook Extra details: Reviews, Acknowledgements, Preface, Cover page, Comments.
- Multimedia Content: PowerPoint presentations, images, Audio, Video.
- Related Resources: exercises, tests, quiz, other supporting materials.

The proposed structure for the meta details is adequate enough to include most of content and info features of most books and content, although further components may be added, if necessary, in the extra details category.

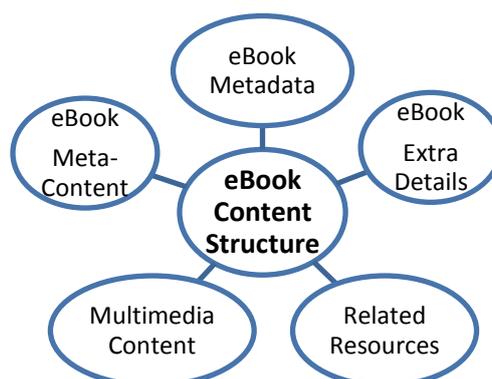


Figure 2. Proposed eBook Content Structure

eBook Meta-Content Building Methodology

Metadata describes the information about content and context of documents or data files, the quality of the original document/file is greatly increased. However meta-content, albeit a broad concept, is anything about a specific content or document such as a book. There can be several sets of meta-content around a piece of content and its logical structure as well as storage, age or access to the content. Meta-content represents a myriad of forms of info and content that can use a database, a filing structure or logic based knowledge representation system for storage and access [15].

The proposed methodology for building meta-content is aimed at capturing, gathering, structuring, and indexing existing or new books content and resources. It is used for improve and efficient usage of finder, sorting content, search engine and organizing information. Furthermore, the digital form of the content will allow, improve and facilitate the following:

- Third party applications to interact more efficiently with the body of the content for an in-depth and precise manipulation of an eBook or eContent,
- Content creation and authoring on the subject by authors, as well as producing with ease other relevant formats such presentations and learning objects.
- Producing different formats and output for wider and richer content publishing.
- Distribution and dissemination through different online and offline education, and other, medium and platforms such as eLearning, virtual learning environment.
- Marketing eBooks, and other related products, with more accurate descriptions and representations of the content.

Fig. 3 illustrates the proposed phases to build the various components of the meta-content, for existing or that will construct a well-structured and digitally accessible content:

- Phase 1: Manual analysis of text and structure of the different components of a Book by the author or/and subject matter expert. The structure can be in terms of chapters, categories, paragraph, topics, themes, meanings, thesaurus, thoughts, or a combination of all the above. This phase is the most important of the whole process because this will structure and dictate the way such content will be collected, manipulated, stored, and digitally accessed for further developments and usage.
- Phase 2: Building, editing or/and extracting the different digital Content components. The different meta-content components are stored in a dedicated database.
- Phase 3: This phase covers the process of proof reading, verification and validation of the digital Content, being entered in the database, with options for comments and suggestions, and reference to a proof reader or a curator.
- Phase 4: Building relationship between the different meta-content components being entered. This will allow for digital content to be built and formatted accordingly, as well as building an eBook digital map for easy access and content manipulation by third party applications. In fact in addition to eBook authoring, dedicated applications can be developed to take advantage of the digital format in order to develop and create different types of eContent such as learning objects, booklets and presentations.

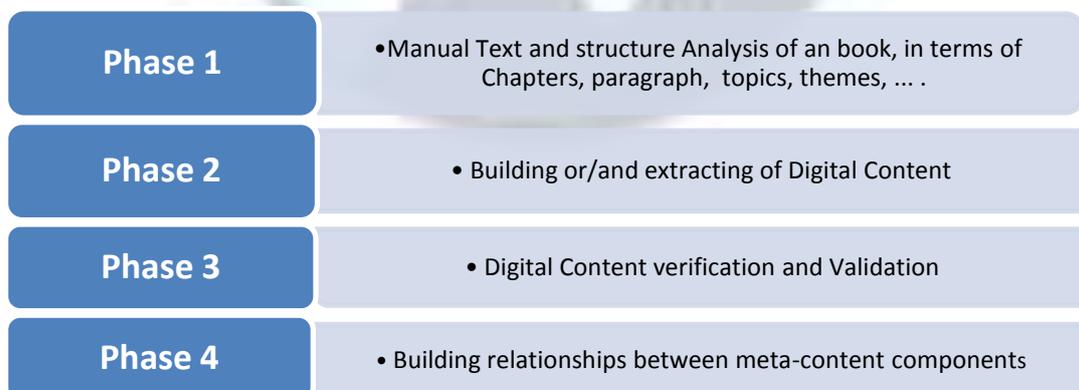


Figure 3. Meta-content building methodology

eBooks Authoring System Design

The proposed Authoring system is a piece of software with dedicated modules that allows a user to easily create an application program with dedicated database, for building eBooks or eContent package, capable of manipulating one or

more multimedia objects. The resulting package consists of various types of content, such as metadata, meta-content, and other related info, that are adequately structured and can be produced in a form of an eBook and/or an e-learning content that is conform to international standard, such as SCORM (Shareable Content Object Reference Model), and other educational and general use formats.

The main system design is presented in Fig. 4, where two main area of design are illustrated: eBook builder and eBook reader. The eBook builder consists of a standard editor with tools for inserting various media types, and the content navigation part that allows authors to select, insert, and go through the different sections of the content, i.e. metadata, meta content, and other related resources. The eBook reader consists of several reader controls and tools for adequately displaying and manipulating the content, and the export format part that enable diverse output of the eBook to be used by users, to be integrated with third party software, or to allow specific manipulation of the content for online usage such as advance search, indexing, and information retrieval. The later formats can be:

- For users: .DOC, .PDF, .PPT presentations, and .EXE application
- For Third party Applications: .XML, .JSON or database.
- For online Apps/Systems: can be any format from the above, SCORM, and in a form of a database or a corpus.

The eBook reader is meant to be an integral part of the eBook application in a form of an exe file, although other options will be provided with the later format to enable to generate export, and authoring the eBook content.

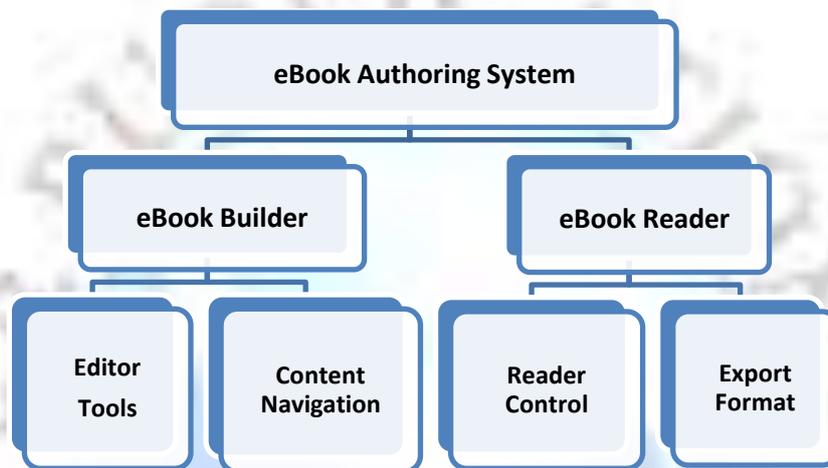


Figure 4. eBook Authoring System Design

Fig 5 illustrates the design of core modules that constitutes the eBook builder authoring system. The multimedia module is a simplified editor that allows for multimedia content and meta details to be inserted/edited. A widely used open source editor has been used for this purpose.

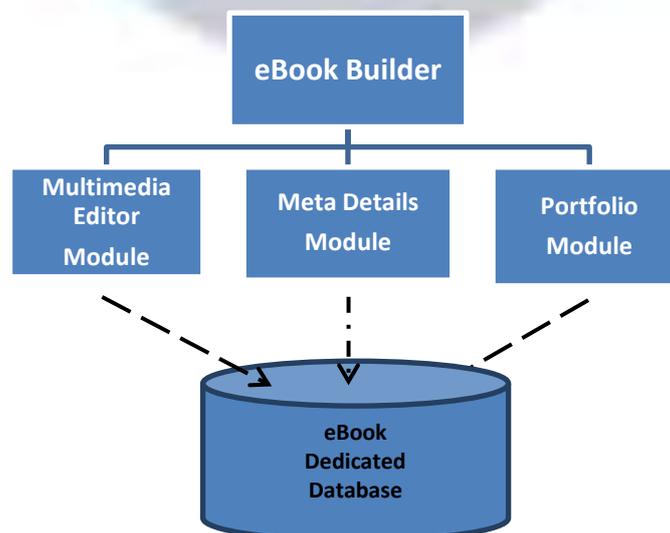


Figure 5. eBook Builder Design

Meta details module has several components such as metadata, meta-content, extra details, multimedia content, and related resources. Each of the components is structured with different and related elements such as chapters and sub-sections for the meta-content. All components and elements that are edited separately are stored in different fields in a dedicated database. This way any book or content is fully digitized and very rich in terms of content and related info. Furthermore, it can be accurately searched and/or manipulated by third party applications, as appropriate, and efficient exploitation of the content can be made.

Panoply of formats can be outputted and exported from the portfolio module. This varies from the way the eBook or eContent is being displayed to materials being produced for online and/or offline delivery and use.

Fig. 6 presents the main modules that constitute the eBook reader. The reader module is a collection of controls and tools that allows the user to display and manipulate the various content of the eBook. Depending on the type of the eBook or documents, i.e. academic book, story book, report... a specific template or theme will be used to display the content. Finally, the export format module enables author or user to select and generate the appropriate format for export. In this case the database is essential an essential part of the eBook reader. The latter can be provided with option with a secure content and a selection of format export.

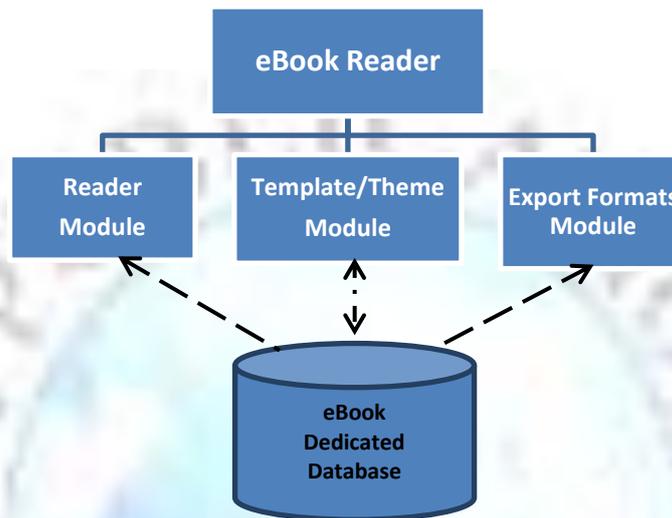


Figure 6. eBook reader design

Design eBook Portfolio

The structure for the eBook portfolio and the various outputs that can be obtained is illustrated in Fig. 7. Through dedicated applications within the authoring system, a rich portfolio of the content is produced which consists of three main deliverable parts: display, product, and export options. The options can be used for online or/and offline purposes.

| eBook Portfolio | | |
|---|--|---|
| Display - eReader - Search - Navigation | Product - eBook - Learning Object - Printed Book | Export - Application - Database - XML, JSON |

Figure 7. eBook portfolio reader design

Once the content and info details are stored in the database, it can be displayed for: reading with an eReader application, advanced content search tool, and navigate through the different content and info using mind-map like features. Products can be produced, as well, such as eBook or eContent document, printed document with specific formats, and learning objects to be used as learning materials for online use.

Finally, the content delivery is facilitated by exporting different formats for efficient distribution, in a form of an application, database, or XML/JSON file. The latter would allow third party software to manipulate the various

components of the eBook content and info, and easily deliver such content through different online/offline medium and platforms such eLearning or Virtual Learning Environment, in a form of eBooks, presentations, and multimedia content.

Implementation and Results

The implementation has been divided into three parts: preparation of the content, development of the eBook builder, and eBook reader authoring system.

For the first phase of the project funded by NOOR Research Centre, some existing books have undergone some processing and analysis for metadata as well as meta-content. Three existing popular books have been selected and their content structure have been prepared, organized, and extracted with clear topics, themes, and categories to be digitized. The process has taken some time because of the manual validation and authentication procedure put in place to make sure of the accuracy and integrity of the content, as well as the way digital content should be stored and indexed. This phase has taken 3 to 4 weeks for each book to be finalized. However, in order to speed up the process an “auto-builder content wizard” for existing books has been developed that would allow the insertion and building of the meta details into a dedicated database. The second phase of implementation of the project is ongoing; the different authoring modules for the eBook builder and Portfolio are being developed as a standalone application as well as to be integrated with a web-based system, such as Content Management System (CMS), Learning Content Management System (LCMS), or a Collaboration System. Fig 8 shows the different parts and panels of the user interface, with structured navigation Panels for authors to build an eBook with related materials, and a repository of different file formats and online resources.

The final phase of the implementation covers the development of functions and tools necessary to build the eBook reader with various options for content display and delivery. This phase is at the analysis and design stage and would be discussed and covered in subsequent publication.

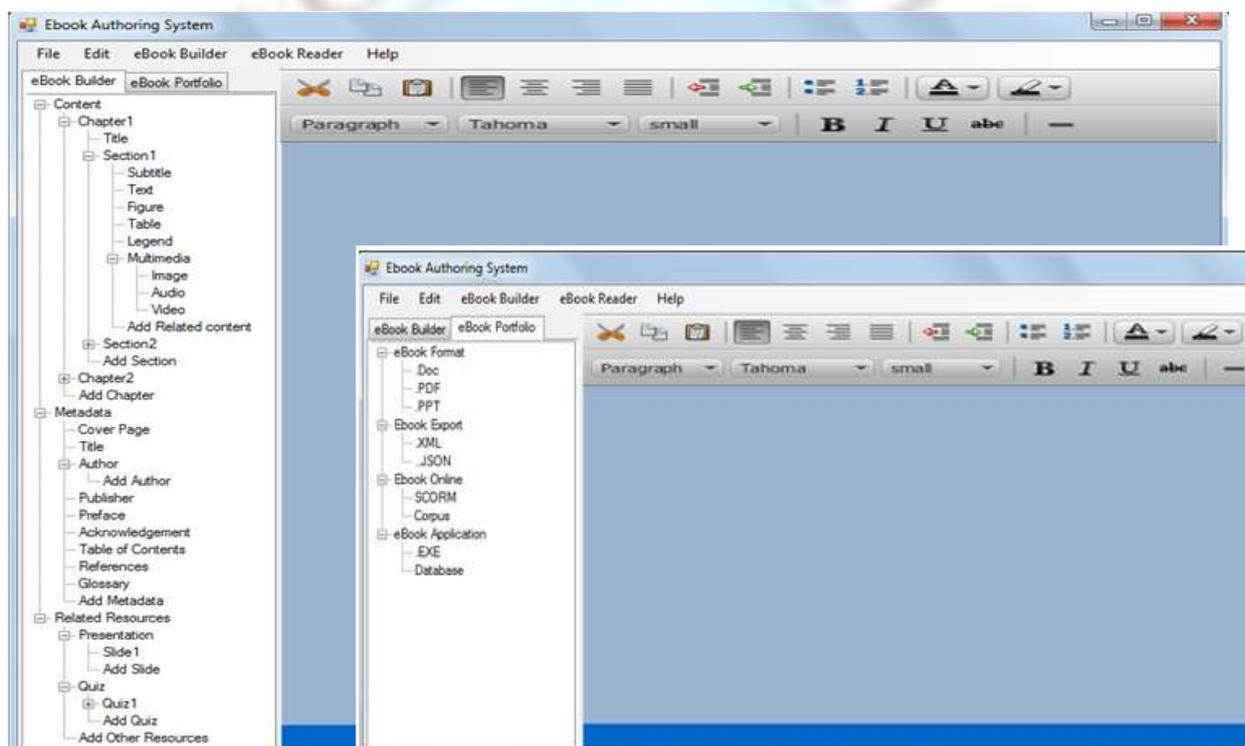


Figure 8. eBook Builder User Interface with Builder and Portfolio navigation panel.

Conclusion/Results

Although the research project is still under development and at its initial stages, overall concepts and design of an authoring system for eBooks and eContent have been presented with the different phases from content analysis to potential content creation and delivery. These authoring system concepts for eBooks and eContent would allow for efficient advanced search, content manipulation and indexing, web content distribution and marketing issues. In addition, advanced search and Natural Language Processing techniques would be implemented more reliable and efficiently for text mining, indexing, tagging, information retrieval of online content and document processing.

In the initial implementation phases, the meta details of three books have been analyzed and extracted into a dedicated database using the “auto-builder content wizard” option. It took some time to finalize the initial phase because of the necessary manual and data validation processes involved. Further development of the system is in progress with more tools being considered for content acquisition, creation and distribution.

Acknowledgment

The Authors would like to acknowledge and thank NOOR Research Centre for their funding, and unconditional support of this research project (Project ID: NRC1-167).

References

- [1]. E.Gardiner, Ronald G. Musto. "The Electronic Book.", The Oxford Companion to the Book, Oxford University Press, 2010.
- [2]. PricewaterhouseCoopers, "Turning the Page: The Future of eBooks ", International Publishers Association, 2010.
- [3]. "List of E-book software", http://en.wikipedia.org/wiki/List_of_E-book_software, last viewed 15 May 2013.
- [4]. Adobe Systems Incorporated , "Adobe® eBook Platform: Authoring and delivering eBooks across devices", Adobe® Digital Publishing, 2010.
- [5]. R.. Andrews, R., C. Haythornthwaite, "Introduction to e-learning research", Handbook of E-Learning Research, 2007.
- [6]. K. A Mills, "A Review of the "Digital Turn" in the New Literacy Studies", Review of Educational Research, June 2010.
- [7]. http://en.wikipedia.org/wiki/Open_eBook, last visited July 2013.
- [8]. <http://ebookarchitects.com/learn-about-ebooks/formats/>, "Details about the most popular eBook formats", last visited October 2013.
- [9]. <http://www.ibm.com/developerworks/xml/tutorials/x-epubtut/>, Liza Daly, "Build a digital book with EPUB", last visited October 2013.
- [10]. http://www.ebookmaestro.com/guide/1_1.html, "eBook Structure", last visited September 2013.
- [11]. <http://www.adobe.com/aboutadobe/pressroom/pressreleases/200910/AdobeandBarnesNobleJoinForcestoStandardizeeBookTechnology.html>, "Barnes & Noble adopts open EPUB eBook Format, PDF and Adobe Content Server", Press release, Adobe Systems. 2009.
- [12]. <http://www.amazon.com/gp/feature.html?docId=1000729511>, "Kindle Format 8 Overview", last visited October 2013.
- [13]. <http://www.mobipocket.com/dev/>, last visited October 2013.
- [14]. <http://www.ecma-international.org/publications/standards/Ecma-388.htm>, ast visited October 2013.
- [15]. R.V.Guha, "Towards a Theory of Meta-Content", http://downlode.org/Etext/MCF/towards_a_theory_of_metacontent.html, last visited 25 May 2013.

6 System Concepts Formulation and Development 129. 6.1 Definitions of Key Terms 129. 6.2 Conceptualization of System Operations 131. 6.3 The System Operations Model 131. 6.4 Formulating and Developing the System Concepts 138. 6.5 Chapter Summary 144. 6.6 Chapter Exercises 145. Part II System Engineering and Development Practices 255. 12 Introduction to System Development Strategies 257. 12.1 Definitions of Key Terms 258. 12.2 Approach to this Chapter 259. 12.3 System Development Workflow Strategy 260. 12.4 Multi-Level Systems Design and Development Strategy 262. 12.5 Chapter Summary 268. 12.6 Chapter Exercises 268.