手机报平台系统要求

一、 分系统具体技术要求

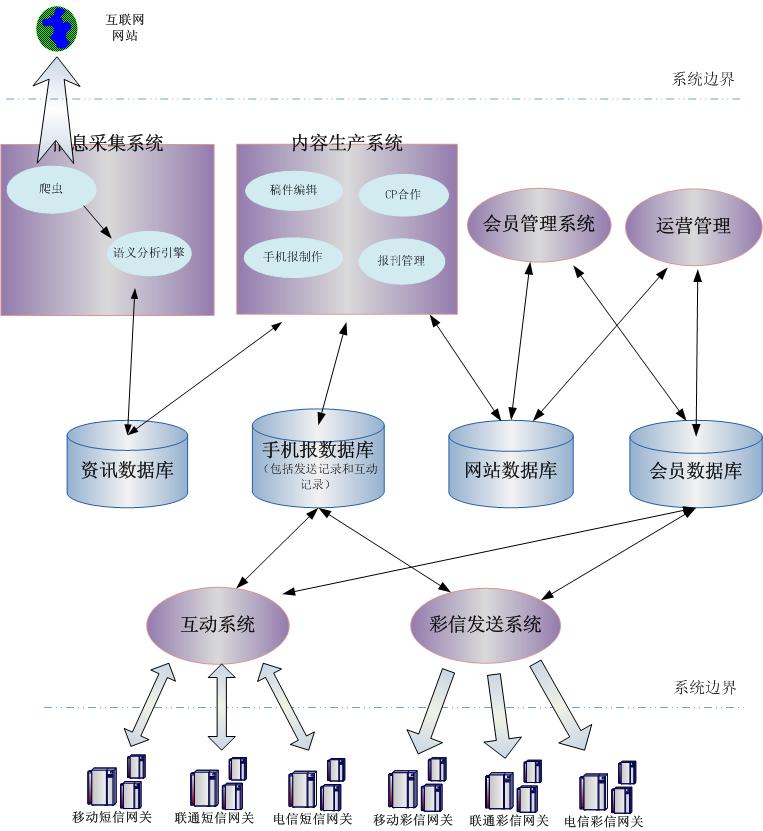
（一）、系统概述

手机报平台是一个以电信、广电运营商合作为基础的，以无线通信信息技术为主导的，以手机教育为特色，以手机等移动通讯设备为接收终端的，集合短信、彩信、IVR、WAP和手机电视等全网无线增值服务的教育资讯传播、学习服务的互动平台。

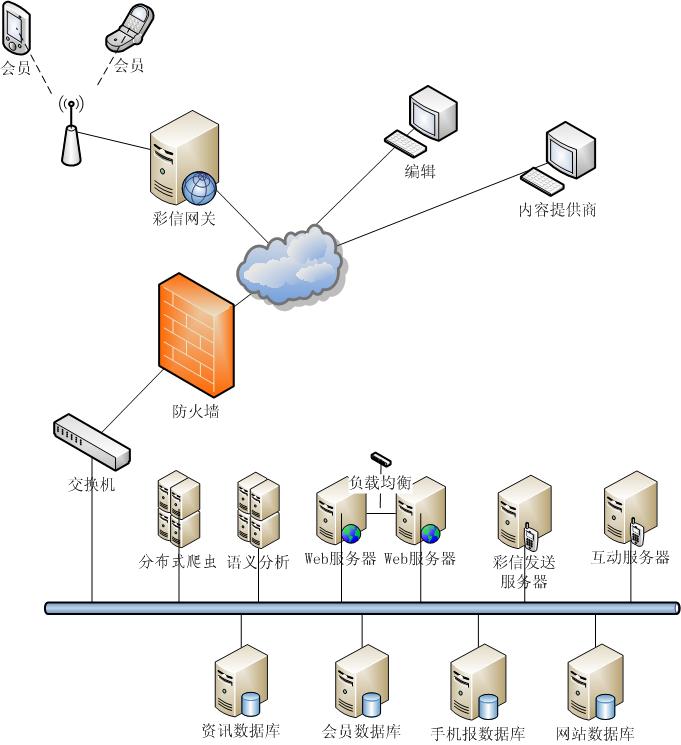
手机报平台要求：支持智能化100份以上的手机报的自动生成；支持千万级的订户发送；支持角色定义及流程管理；支持多机数据备份；至少在线保留最近180天的内容素材和手机报发送记录；支持手机报发送中的停止；支持手机报发送成功与否的状态查询和统计；支持独立防火墙及独立入侵防御；支持多种主流数据库系统Oracle、DB2、Informix、Sybase、MS SQL Server；支持病毒入侵防御及远程病毒清除；

**系统构架**

**网络拓扑图**



**系统结构图**



（二）、系统功能

**手机报定制****内容生产平台**

内容生产平台包括手机报生产中的资讯初编、加工、彩信编辑、发送几个环节，可以支持并扩展到将来为Wap网站、手机客户端管理内容。本平台最大限度的考虑减少人的工作量，通过内容智能分析采集精准的内容，设计合理的人员角色和工作流程，加快手机报的编辑制作。

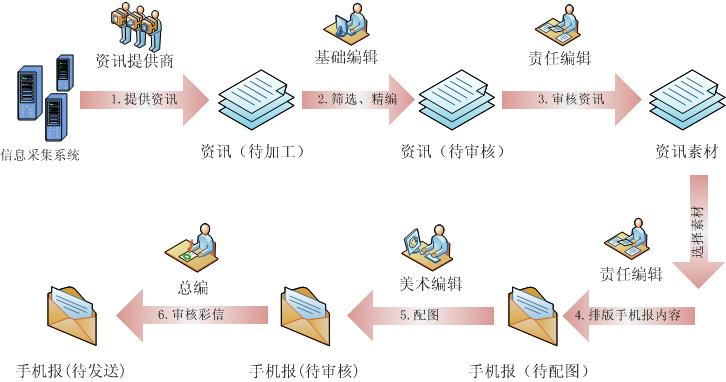
1. 系统特点

* 内容智能分析，分析出内容是否符合杂志定位，自动对内容进行归类分级。
* 加入工作流，流水线作业生产手机报。
* 支持模板，模板化固定的内容。
* 可以按目标用户所在城市，生成包含相应城市特定内容的彩信，例如一些地方资讯。
* 信息自动采集，内容智能分析，精准的筛选出符合教育手机报定位的素材。
* 基于B/S结构，突出教育、手机报系统的运营特点。
* 方便的操作流程，快捷生产手机报。

1. 角色定义

|  |  |  |
| --- | --- | --- |
| 角色名称 | 职责 | 用到的功能模块 |
| 内容提供商 | 每日提供一定数量的资讯。 | 资讯初编、资讯统计、定制帧制作 |
| 基础编辑 | 1. 每个基础编辑会负责几个分类栏目。 2. 对责任分类下的资讯进行初编：   对采集来的资讯（或素材）和内容提供商提供的资讯进行筛选和加工，筛选出手机杂志需要的资讯，并精简编撰成适合手机彩信的内容。   1. 录入素材。 | 资讯初编、资讯统计 |
| 责任编辑 | 1. 对基础编辑处理过的资讯进行审核。 2. 责任编辑负责特定的手机报，为手机报选择内容并排版。 | 资讯审核、杂志制作，  责任编辑可以使用基础编辑的所有功能。 |
| 美编 | 为手机报内容制作插图，生成最终彩信。 | 杂志制作 |
| 总编 | 1. 对美编制作好的手机报进行最终审核，只有审核通过的才能发送。 2. 管理编辑和责任编辑帐户，分配责任栏目等。 3. 统计查询资讯，分析业务情况。 | 杂志审核、帐户管理、资讯查询，管理杂志  总编可以使用所有角色的功能。 |

1. 工作流程



内容生产平台的工作流程如上图所示。

第一步：内容提供商和信息采集系统为平台提供资讯。

第二步：基础编辑对采集的资讯进行筛选，剔除没用的资讯，对留下来的资讯进行加工，缩短标题，把正文精简成适合彩信的小段文本，可能还要修改资讯的分类。

第三步：责任编辑审核基础编辑加工过的资讯，审核通过的进入资讯素材库，审核不过的，填写原因，返回给基础编辑继续修改。

第四步：责任编辑从资讯素材库中选择资讯组成彩信各帧内容，然后提交给美术编辑配插图。

第五步：美术编辑对责任编辑提交的彩信配插图，必要时可以调整彩信的帧，编辑好后提交给总编审核。

第六步：总编审核美编后的手机报，审核通过后的手机报不能再修改了，进入待发送状态，在到达杂志设定的发送时间时，自动发送给订阅用户。审核不通过的手机报，填写原因，回退给责任编辑。

1. 功能描述

（一）资源聚合平台

资源聚合平台提供的并不是一个简单的爬虫解决方案，而是“智能语义聚合应用框架”，它把网络抓取任务和语义解析任务合二为一，可以随不同的垂直领域应用需求而变化。

它分为两个子系统：分布式爬虫和语义引擎。

分布式爬虫

1.1分布式Client-Server架构

采用分布式Client-Sever架构，如果一台连接到互联网的机器是可信任的，那么就可以作为分布式爬虫系统的客户端（Client），他可以通过从服务器（Server）领取任务来进行内容抓取；这样做的优点是，内容抓取速度快，不易被对方网站屏蔽请求。

参考hadoop的分布式实现，包含以下六步的完整任务：

列表页抓取，或关键入口数据抓取；

目标页面抓取；

语义处理；

商业逻辑处理；

数据入库；

更新缓存或本地文件等操作。

分解为一个一个的单独的任务，定义好上下文，让不同机器的客户端去处理，并且可以处理各种分布式计算带来的问题，如领取任务正常，返回任务处理结果时超时，需要收回此任务，重新分配；如客户端失去响应，等等。

1.2提取模板定义

网页内容的精确提取，爬虫通过正则表达式匹配和其他的技术手段，精确提取网页中需要的数据；不同网站的定制逻辑简单明确，只需要指定具体的参数即可，后续的维护也很简单。

1.3链接去重

如果一篇文章的链接已经访问过，并且该文章已通过引擎计算和数据入库，那么在后续的数据抓取中，将跳过此链接地址。

1.4任务分发中心的分布式处理

Server后台也采用分布式架构，足够应付大数据量的处理。

1.5性能

采用流行脚本语言 Python 构建的程序，简单易读，而且由于 Python 的底层是C语言，所以性能与C语言构建的爬虫程序不相上下。

语义引擎

语义引擎能快速处理以下语义任务：

分词、分类、去重、实体识别、提取标签、情感计算、文本相似性

与分布式爬虫的任务分发中心相结合，可以达到实时或近似实时计算的效果。

2.1去重引擎的主要特点及技术

特征提取选择采用 Shingle 方法,并且抛弃无意义的片断，节省存储空间;

进行编码压缩,利用32bit MD5 HASH方法对特征进行编码进行压缩;

建立特征索引表，加快计算重复文档速度;

主要优点，去重正确率较高，计算速度快，操作方便简单。

2.2自动提取文章标签的主要特点及技术

自动分词，并标注词语词性，从而识别出实体词语;

区分标题中词语和文章正文内容权重，加重标题词语权重;

通过相应训练语料，训练出IDF权重大的词语作为停用词，以防止标签提取的不准确性;

针对实体词语的权重排序，并过滤掉停用词作为最后的标签;

主要优点，计算速度快，提取标签正确率高，可以人工添加停用词对标签提取进行人工优化。

（二）资源统计生产平台

A、资讯统计

按照资讯的采集时间、状态、来源、所示分类查询统计资讯。

* 内容提供商、基础编辑、责任编辑、总编都可以使用本功能。不同角色，有不同的数据访问权限。
* 内容提供商只能统计自己提供的资讯。
* 基础编辑只能统计自己处理过的资讯。
* 责任编辑只能统计自己负责的分类的资讯。
* 总编可以统计所有的资讯，总编除了按时间，分类查询外，还可以按下面两种方式查询：
  1. 按基础编辑查询：查询初编后的，且初编时的基础编辑帐户是选择框选中的帐户的资讯。
  2. 按责任编辑查询：查询审核过的且审核人是选择框选中的责任编辑的资讯。

B、初编

初编是指对资讯素材的修改加工，是内容提供商录入资讯和基础编辑提取资讯进行编辑的功能。按角色不同操作方式不同：

* 基础编辑
  1. 提取自己负责分类下的待编辑资讯并锁定，避免两个编辑同时编辑一篇资讯。
  2. 编辑提取的资讯，编辑短标题和摘要，缩减为适合彩信的长度。
  3. 支持富文本方式编辑。
  4. 如果资讯的分类不正确，还可以修改分类。
  5. 为地方性资讯设置省市。
  6. 处理被责任编辑驳回的资讯。
  7. 删除没用的资讯
  8. 优先提取优先级高的资讯。
  9. 录入资讯，录入标题、正文，为资讯上传配图，设置资讯分类。
* 内容提供商
  1. 录入资讯。
  2. 删除自己录入的没用的资讯

C、资讯查询

此功能是为了在有很多资讯时，方便处理紧急新闻而设计的。只有总编可以使用本功能。

* 根据时间、标题包含的关键字、状态、来源和分类几个条件查询资讯
* 可以修改查询到的资讯的优先级和分类，优先级高的将被初编优先提取处理。
* 可以把查询到的资讯指派给某个基础编辑马上处理。
* 可以删除查询到的资讯。

D、资讯审核

* 列出带审核的资讯，责任编辑只能审核自己负责分类下的资讯。
* 审核基础编辑加工后的资讯，检查内容是否精简、通顺、完整，标题是否精简，分类是否正确。审核过程中可以查看资讯原文。
* 审核过的资讯可以撤销和再驳回。
* 责任编辑和总编可以使用本功能。

E、杂志列表

* 列出当前制作完的杂志，包括已经发布的和正在编辑中的。责任编辑和美术编辑都在这个画面中制作杂志。
* 可以按照杂志的类型、状态、标题中包含的关键字查找杂志。
* 可以删除未发送的杂志。
* 查看或者修改杂志的预定发送时间。
* 查看或者修改杂志的预定发送用户，可以按照订阅用户、地域、年龄、性别、卡类型等从用户库中提取用户进行发送。

F、杂志制作

* 制作彩信杂志，设置制作的杂志标题（期号），动态增加帧或者删除帧，为帧设置图片、文字、声音和播放时间。
* 可以为帧设置受众城市，只有发送给目标省或者城市的用户的彩信中才会出现本帧内容
* 编辑过程中可以直接预览帧的内容。
* 提供素材选择区，可以从审核过的素材中直接选取素材，自动加为帧的内容。
* 提供常用标记（固定短语）选择区，可以在帧正文中插入常用的标记。
* 编辑过程中可以试发送彩信到手机，查看实际效果。
* 责任编辑制作的杂志需要提交给美编处理，美编为文字内容配插图，然后把杂志提交给总编进行终审。

G、定制帧制作

这是内容提供商用来制作手机报中定制帧的功能。定制帧的说明见4.11杂志管理。

按照编辑彩信帧的方式，为手机报的定制栏目帧编辑内容，可以根据情况动态地为栏目增加帧。

H、杂志审核

* 总编审核最终的杂志，杂志只有审核通过才能发送。
* 列表方式查看待审核的资讯。
* 查看杂志每帧的内容，检查内容是否合法，选材是否精准。
* 审核时可以试发送彩信到手机，查看实际效果。
* 审核不通过的，填写驳回原因，把杂志驳回给责任编辑。

I、修改密码

修改自己的帐户密码。

J、帐户管理

* 管理内容生产系统的帐户，进行增加、修改、删除。
* 为帐户赋予角色，为帐户分配负责的资讯分类。
* 限制只有总编可以使用本功能。

K、杂志管理

* 管理杂志（例如早期教育报、中学生报），增加、修改、删除杂志。杂志的基本信息有杂志编号、杂志名称、发送周期、发送时间、对应的资讯分类、简介、受众群体。
* 管理杂志模板。

杂志中可能有一些固定不变的内容，例如开头加日期和天气预报，最后一页是杂志封面，倒数第二页是订阅方式等，中间一些固定栏目的标题和格式等，这些不变的内容做成模板帧，把模板帧按顺序组织起来形成杂志的模板，这样在创建杂志时，按照模板自动填充内容，方便处理。

* 帧模板管理
  1. 建立、修改，删除模板帧。
  2. 编辑模板帧的方式与编辑杂志的一个帧相似。
  3. 模板帧的信息有模板名称、文字、图片、声音、播放时间、受众城市、帧的位置，是否自动启用帧。自动启用的帧是指在创建一期杂志时，自动插入到彩信中的帧。
  4. 定制模板帧：按照受众属性和内容来源两个特征为条件来组织内容的模板帧，可以组合使用多个条件，类似邮件过滤规则，在发送手机报时，按照过滤条件，为不同属性的会员生成不同的帧内容。

受众属性从下面几个条件中选择，可以多选：

所在城市、性别、年龄、所属机构、职业、爱好、教育程度、收入水平。

内容来源从各内容提供商（CP）中选择，只能单选。

例外规则：当受众不在各条件范围中时，适用的规则。

定制模板帧示例：例如在“大学生就业报”的校园新闻栏目帧中，设置四个规则：1. 受众目标为北京大学的学生，使用CP北京大学提供的内容；2. 受众目标为清华大学的学生，使用CP清华大学提供的内容；3. 受众目标为传媒大学的学生，使用CP传媒大学提供的内容；4.受众目标不在前面三个条件范围内的话，使用教育台提供的内容。那么，在发送手机报时，发给清华大学学生的彩信中，这个帧的内容就是清华大学提供的内容，发送给传媒大学学生的彩信中，这个帧的内容就是清华大学提供的内容，发送给人民大学学生的彩信中，这个帧的内容就是教育台提供的内容。

* 设置添加素材时的格式

添加素材时的格式是指向帧中添加资讯等文字素材时，自动应用的格式，系统支持四个标记：素材的标题，正文，来源和作者。

* 管理杂志常用标记

常用标记是一些常用的固定短语，编辑手机报时可以直接拿来使用，标记分为全局适用和本杂志适用两种，全局适用的就是在编辑各杂志都可以使用，本杂志适用的就只有本杂志可以使用。

* 1. 增加、修改、删除常用标记。常用标记的信息有名称、内容和适用范围。
  2. 支持日期和天气预报两个固定标记，在帧的文字中插入了这两个标记，在发送彩信时，要把日期替换成发送时的日期，天气预报替换成目标号码所属城市的天气预报。

**会员管理**

会员系统的使用者可以分为运营、渠道、客服三种。运营人员可以使用所有功能，渠道推广人员可以导入会员、管理自己发展的会员和集团会员。客服人员可以查询会员，修改会员信息。

1. 会员管理

* 按照手机号码、姓名、性别、订阅杂志、城市、年龄、所属机构、职业、爱好、教育程度、收入水平、活跃度等条件查找会员，查看会员信息和互动记录。
* 修改会员基本信息，包括姓名、手机号码、性别、年龄、城市、职业、爱好、教育程度、收入水平。
* 修改会员订阅的杂志和认证状态（是否许可）。认证状态就是指用户是否许可了可以向他的手机发送手机杂志。
* 修改会员的计费方式和有效期，在有效期内的会员才能接收手机杂志。
* 会员活跃度，根据会员参与短信互动、热线电话的情况，确定会员的活跃度。

1. 批量导入

* 批量导入会员数据，通过特定格式的excel文件导入会员数据。Excel包括的用户属性有手机号码、姓名、性别、年龄、职业。
* 导入时可以设置会员数据分类属性和用户属性。

数据分类属性：会员的认证状态、认证方式、所属来源（推广渠道）、所属机构、订阅的杂志。

认证方式：就是发展用户的方式，有上行注册、内部会员、销售体验、电话外呼，可以增加认证方式。

所属来源：指发展用户的渠道，例如教育报合作，可以增加新的渠道来源。

所属机构：指用户所属的机构，对于集团会员就是所在的集团，例如清华大学。

用户属性：年龄段、性别、区域、城市、地址、兴趣、教育程度、收入水平或其他。

* 批量信息修改：批量导入excel中的用户信息，把会员库中已经存在的会员信息用excel中的信息替换。业务人员在批量导入时有时会把数据搞错，使用本功能批量纠正错误的数据。

1. 查询统计及报表

* 按周、按月、按季度、按年为周期，统计各认证方式发展的会员数量，统计订阅各杂志的会员数量。
* 统计从各机构来的会员的情况。
* 统计结果可以输出为excel报表，可以以饼图、曲线图方式直观查看数据。

1. 集团会员管理

集团会员是指以组织机构或者单位的身份来订阅杂志，向组织成员发送杂志，最后由组织机构付费的组织机构。例如：某所中学由学校来付费，为学生免费订阅“中学生报”，这个中学就是我们的集团会员。

* 集团会员的添加、删除，信息修改。集团会员信息有名称、地址、联系人、联系电话、订阅的杂志、会员规模、所属来源（发展渠道）、计费方式、有效期。
* 集团会员查询。

**互动管理**

1. 上行订阅\退订杂志

短信上下行系统，会员可以发送短信订阅杂志或者退订杂志。

1. 短信互动

自定义上行命令和下行的短信内容，实现用户上行一条短信，保存用户上行内容，并自动回复一条短信。

可以根据号码和时间查询用户上行短信记录。

**彩信发送系统**

1. 发送系统

与彩信网关通讯发送彩信杂志，在杂志的预定发送时间，把彩信发送给订阅用户和选择的号码（按照地域、性别等特征选择的一定数量号码）。

* 支持多个发送通道，可以设置当前默认使用的通道，可以根据手机号段选用不同通道。
* 支持黑、白名单，不能向黑名单中的号码发送杂志。如果启用了白名单，只能向白名单中的号码发送杂志。
* 网关故障报警，在彩信网关发生故障时，通过短信形式向管理员报警。
* 失败重发，对于发送失败的杂志，尝试3次重发。

1. 计费系统

* 计费方式为包月计费，计费主体是个人会员和集团会员，可以有免费会员。
* 生成个人会员和集团会员的月计费对账单。

**运营管理**

1. 配置管理

管理系统的配置参数，

* 设置是否启用白名单
* 设置杂志发送的时间段，例如只在8点-22点之间发送杂志。
* 设置接收报警信息的手机号码。

1. 字典管理

* 管理素材分类字典。
* 管理手机号段字典，支持号段数据导入。
* 管理手机厂商、手机价位，收入水平、职业等会员属性字典。

1. 黑名单管理

黑名单就是拒绝接收彩信的手机号码名单，黑名单管理就是对这些号码的增加和删除。

可以根据手机号码在名单中查找号码。

1. 白名单管理

白名单就是可以接收彩信的手机号码名单，白名单管理就是对这些号码的增加和删除。

可以根据手机号码在名单中查找号码。

1. 彩信发送记录查询

按照发送时间，手机号码等条件，查询彩信发送记录。

按照发送时间，杂志期号，查询发送失败的记录。

**数据备份**

安全、完善的数据备份，支持灾难恢复和事务日志回滚，支持联机热备份。需要备份的数据主要有：软件系统，会员信息，内容素材，手机报期刊和发送记录。至少在线保留最近180天的内容素材和手机报发送记录。

**系统安全性**

采用标准的Web安全技术，防止暴力猜密码、Sql注入攻击等。系统支持SSL、IPSec等。

关键数据可以考虑加密存储。

采用防火墙隔离内外网，采用入侵检测系统等保证网络系统的安全。

**客户端系统**

客户端的开发基于上述手机报方式，是此系统的升级完善版本，除保留上述手机报系统的现有功能外，更可以实现更加便捷与高效的下发、浏览、管理等功能，并成为教育电视台资讯增值互动的主要支撑平台。

以下从功能实现角度对客户端进行描述：

**（一）客户端功能需求**

* 将用户手机与学习超市网站注册之用户名绑定；
* 基于无线互联网，建立自有的IP PUSH平台，为学习超市的用户提供手机；
* 统一消息服务，并能够实现用户从PC收发消息；
* 操作过程有与短信完全一样的用户体验，用户收到信息时，得到象短信一样的提醒。发送信息时，只需要选择使用客户端即可；
* 同一条信息，可以发给手机号码、邮件地址、学习超市网站。支持抄送、密送、填写主题、可以添加图片、铃声等各种附件；
* 收到的消息可以转发到手机、邮件或者学习超市；
* 通过PC机发短信，支持三大运营商。无需添加好友，且对方看到的发送号码是自己的手机号；
* 电脑收短信可回复。收到的短信与邮件在同一个收件箱管理。可进行各种排序，搜索，可建文件夹，管理更方便；
* 群发、定时发送；
* 从网页可以同步、备份、下载通讯录；
* 从手机也可以发起操作；
* 通讯录可以建立群组，群发信息更方便；
* 可以从手机上传图片、博客到网站；
* 好友在网站的更新、私信可以在手机上收到，可立即回复、评论等；

**（二）客户端下载、安装、使用**

* 登陆学习超市后，推荐用户下载，并与账户绑定实现计费；
* 对于安装客户端的用户，可现实不包月定制增值业务，即可收发各类资讯；

**3G应用**

随着资讯下发、网站互动的铺开与深入，教育台的手机增值服务会进一步拓展与开发，进军3G背景下的手机增值业务空间。

**建立WAP站点**

WAP是一种无线应用通讯协议(Wireless Application Protocol，简称wap，是一个全球性的开放协议，是移动通信与互联网结合的第一阶段性产物。这项技术让使用者可以用手机之类的无线装置上网，透过小型屏幕遨游在网站之中。

WAP站点可承载大量的文字、图片及多媒体信息。建立教育台自主的WAP站点，于教育台而言，一方面，使得无线互联网的优势在学习超市的用户群中得以展现，另一方面，将学习超市网站模式移植到无线互联网，以获得更大基数的用户群体。于用户而言，是为学习超市的目标用户提供了随时随地的、多层面、多途径的互动途径。具体可以开展的业务如下：

* 包括集团与个人用户的文字资料的传播与分享——平台的编辑、上传等功能已相当成熟；
* 各类相关视频的下载、点播；
* 无线互联网模式下的社区互动、群组管理；
* 实现邮箱、博客、游戏等传统互联网功能。

**开放IVVR业务**

IVVR（Interactive Voice and Video Response）即交互式语音及视频应答。IVVR是一种全新的无线语音及视频应答增值服务，视频电话用户通过拨打指定号码，获得所需信息或者参与互动式的服务。IVVR借助视频和语音的特色，突出“互动”型功能。此功能在教育台的应用十分广泛，如下：

* 各类教育视频资源的点播、收看；
* 与系统的交互——实现可视条件下有问有答的互动；
* 视频短信、视频内容点送；
* 视频呼叫中心——可通过视频表达的丰富性来展现更多的内容。方便用户获取更多信息。

**运营手机视频**

手机视频在现有网络环境下，已经可供手机用户使用。3G背景下，手机视频的观赏效果将更加流畅、画质更加清晰。

* 建立手机视频垂直频道——教育频道，将视频文件集成在频道资源，方便用户选择；
* 上传视频的展示专区——用户自主资讯经审查合格后，挑选精品在手机视频频道中展现，供其他用户下载、点播；

实现稀缺教育资源的在线直播，方便所有学习超市的用户在各种环境下实时接收资讯，使得更多的用户分享资讯。