



核电企业内容分发引擎的设计与应用

汪鹏, 涂红兵, 王云福

(深圳中广核工程设计有限公司, 广东 深圳 518000)

摘要: 内容分发是核电企业的重要业务, 负责在公司、部门、项目组等层级之间传递结构化/非结构化的内容媒体。内容媒体结构复杂、参与分发层级众多、分发业务场景多变, 给分发引擎的设计带来很大挑战。经过对公司分发业务的梳理, 设计一种可动态配置的内容分发引擎, 该分发引擎支持业务逻辑配置化, 减少了由分发业务变化导致的引擎频繁变更。同时结合数据库触发器技术, 设计了一种引擎前期的内容抓取机制, 可大大降低待分发内容的抓取时间复杂度和空间复杂度, 保证了任务产生的即时性。

关键词: 内容分发; 动态配置; 内容抓取; 即时性

中图分类号: TM623

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019033

Design and application of content distribution engine for the nuclear power enterprise

WANG Peng, TU Hongbing, WANG Yunfu

China Nuclear Power Design Co., Ltd., Shenzhen 518000, China

Abstract: Content distribution is an important business of nuclear power enterprises, and is responsible for transferring structured/unstructured content media between companies, departments, project groups and other levels. The complexity of the content media structure, the participation in the distribution hierarchy, and the variety of distribution business scenarios pose great challenges on the design of the distribution engine. After sorting out the company's distribution business, a dynamically configurable content distribution engine was designed, which supported business logic configuration and reduced frequent engine changes caused by changes in distribution business. Combined with the database trigger technology, a pre-engine content crawling mechanism was designed, which could greatly reduce the crawling time complexity and space complexity of the content to be distributed, and ensure the immediacy of the task.

Key words: content distribution, dynamic configuration, content-catching, extemporaneity

1 引言

随着第三代核电技术的逐步推广, 在核电站

建设过程中出现了数量越来越多、种类越来越复杂的内容媒体, 主要包括工程信函、技术文件、竣工资料、工程照片及视频文档等。这些内容媒



体数据是核电企业最为重要的资产之一，对于企业的发展以及创新有着重要的意义，在企业内容媒体生产完成之后，需要根据媒体内容进行内容的逐层分发与办理。核电企业通常都会将此类数据归档至企业内容管理平台中，通过企业内容分发平台进行统一的逐级分发、管理及利用。核电企业内容复杂多变，从设计到实体产生再到现场施工，其间与众多的集团内外部公司、部门、项目组均存在接口关系，业主、项目部、工程总部、设计院、第三方公司等每个公司或部门的分发方式与分发类型均存在差异。如何建立高通用性、高扩展性和高可靠性的分发引擎，成了企业内容分发平台所面临的一大挑战。

2 研究背景及现状

2.1 研究背景

目前核电项目呈现多项目基地、多条技术路线特征，参与专业众多，涉及工程设计、设备制造、现场施工、调试运行等方方面面，同时分包院、承包商、供应商等参与公司众多，接口十分复杂。为了保证内容媒体在生产之后能够及时、准确地分发传递到各相关部门，核电企业一般都设置了专门的文档管理部门，作为内容的归口管理部门，负责内容媒体从产生、整理、保存、归档到分发利用的整个流程管理。这种“内容媒体全过程一体化管理”模式，为核电项目顺利推进提供了有力而规范的保障。

2.2 研究现状

核电工程文档管理平台（AED）是中广核工程有限公司建立的企业内容统一管理 & 分发平台，全面承载了公司核电工程内容媒体的生产和分发，是公司最为重要的内容管理平台和数据中心。根据前后台业务类型和业务团队设置及管理模式的差异，核电项目设计模块和工程模块对内容媒体的分发方式存在差异。设计部门一般以设计总工程师（简称设总）团队为中心，由项目设

总根据媒体内容来进行分发；工程部门则由函件分发委员会（Correspondence Assignment Committee, CAC）依据分发清单实施统一分发，图 1 展示了前后台的分发模式。

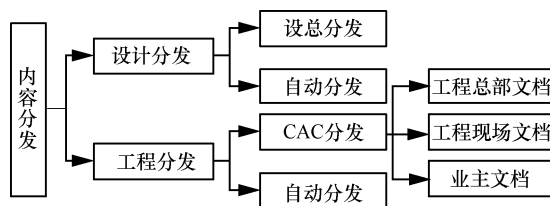


图 1 AED 分发模式

内部内容媒体生产完成之后，会归档到核电企业内容管理平台，外部来函则由文档人员手工录入内容管理平台，统一由文档人员在内容管理平台中补充分发信息，内容分发引擎根据分发信息启动分发，全过程如图 2 所示。内容分发平台自动抓取待分发数据，由分发引擎根据内容的发送方和接收方信息，自动启动第一次分发，给设总、CAC 和普通用户产生待办任务，随后由其自行根据业务内容进行后续逐层分发。

由于内容媒体（主要指来函、发函）的生产涉及的发送方与接收方众多，且每个发送方/接收方的分发方式、分发类型和接收对象均不同，导致了分发引擎需要维护大量的分发类型与分发方案，来启动第一次分发。随着项目的推进，这些分发方案随时可能发生变化。设计院作为发送方、其他公司/部门作为接收方时的分发情形如图 3 所示，其他公司为发送方的情形与此类似。在进行内容分发引擎的设计时，需考虑不同公司互发的所有情况，直接采用硬编码的形式，会给内容分发引擎带来很高的复杂度，而且无法应对分发业务的频繁变更，通用性、可靠性和可扩展性比较差。本文提出了一种可配置化的分发引擎设计方案，将业务模拟成一条分发规则，利用分发规则计算算法自动计算出内容媒体的分发类型，实现一次分发。

本文相关术语解释见表 1。

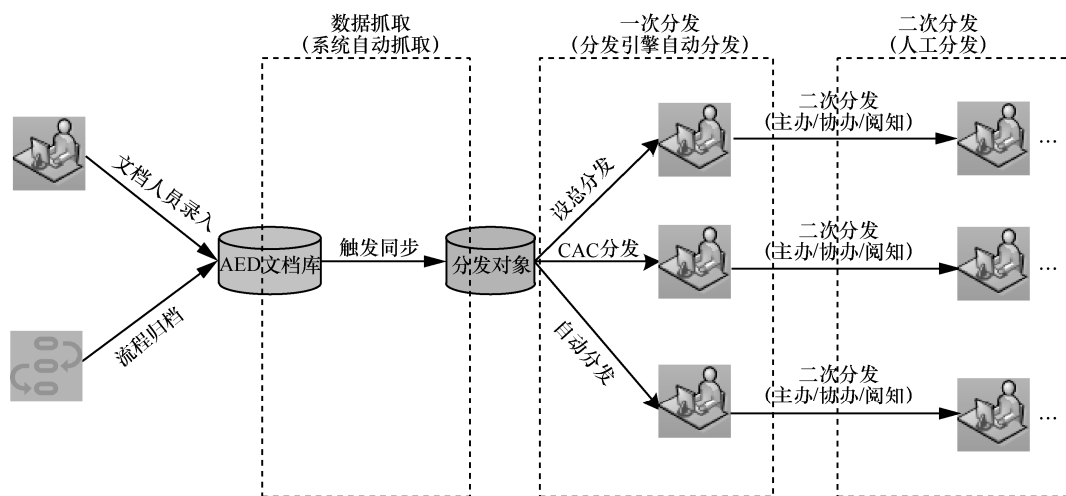


图2 分发流程

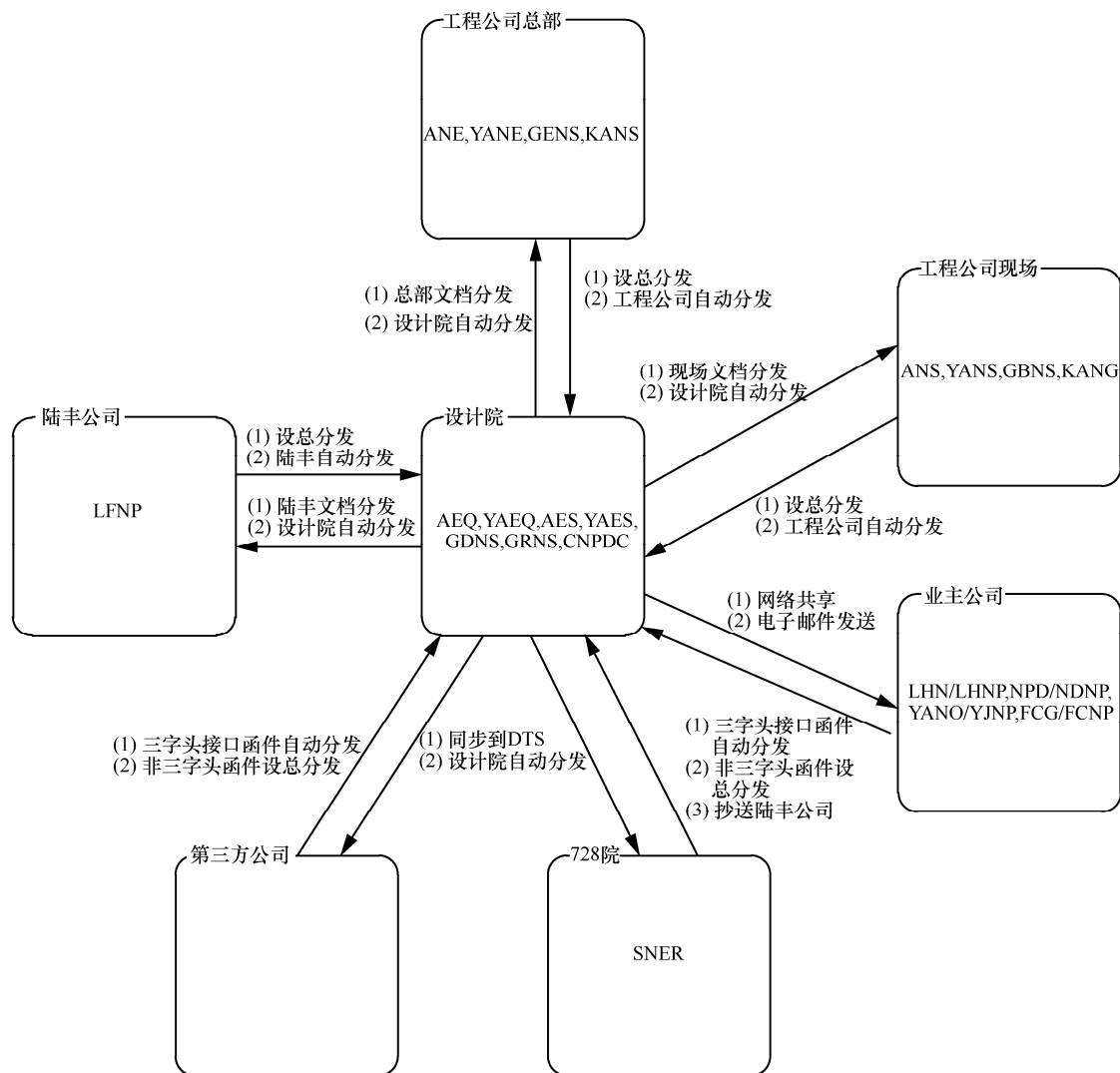


图3 设计院作为发送方时的分发情形



表 1 相关术语解释

序号	术语	说明
1	内容媒体	工程信函、技术文件、竣工资料、工程照片及视频文档等，为方便管理，内容媒体根据发送方/接收方等信息进行了统一渠道编码
2	发送方/接收方	内容媒体产生和接收的渠道标识，一般用三字码或四字码标识
3	一次分发	由内容管理平台通过内容分发平台依照内容的分发信息自动进行第一层级的分发，可分发到设总、文档、用户个人、项目组等
4	二次分发	设总/文档分发到普通用户个人，或者普通用户个人之间的分发
5	设总分发	设计团队的设计总工程师进行任务接收/分发
6	CAC 分发	函件分发委员会进行任务接收/分发
7	内部发送/接收	系统内公司间互发，称为内部发送/接收
8	外部发送/接收	系统内公司与外部公司之间互发，称为外部发送/接收
9	分发信息	内容媒体的属性字段，主要包括主办、协办、阅知、外部抄送、内部抄送等，表示了内容媒体的分发信息

3 内容分发引擎设计

内容分发引擎的设计需考虑以下 3 点。

(1) 即时性

核电内容管理平台作为基础数据中心，存储了海量的核电内容媒体数据，需快速抓取出待分发的数据，保证引擎分发的即时性。

(2) 可扩展性

核电企业内容分发平台承载了多家参与方之间的分发业务，每家公司作为发送方和接收方时的分发和接收方式都不同，而且随着项目进度和质量的要求变化，分发方式可能发生变化，分发引擎必须能灵活地应对此种情况。

(3) 通用性

核电项目参与方众多，但设计类和工程类参与方模式大体一致，对于新的参与方的加入，分发引擎应做到“无缝衔接”，满足新的参与方的分发业务需求。

本文设计的可配置化内容分发引擎方案，可以解决以上 3 个问题。

3.1 引擎结构设计

本文提出的可配置化内容分发引擎方案，将每项分发业务抽象成分发规则，放入规则库。分发业务发生变更时，只需调整规则库中对应的分

发规则即可，无需在代码层面做任何变更。同时，设计了高效的分发类型计算算法，可在 $O(N)$ 的时间复杂度内根据发送方、接收方和分发规则，计算出一条内容媒体的分发类型。分发引擎结构如图 4 所示。

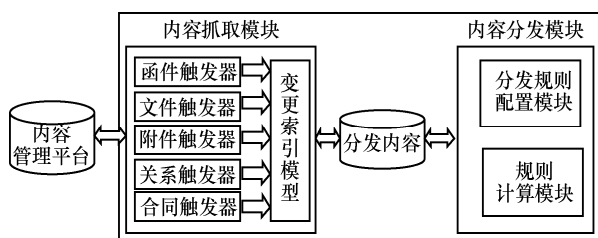


图 1 内容分发引擎结构

各个功能模块说明见表 2。

3.2 内容抓取

内容管理平台（ECM）用于存储基础业务数据，具体包括工程函件、技术文件、附件、合同、函件—文件/附件关系、通信渠道号、项目代码等信息。ECM 作为基础库，存储了大量的基础数据，对于创建时间久远、已经分发的、标志删除的、不需分发的数据，必须设计一种抓取机制，能够实时、快速地从这些海量基础数据中提取出待分发的数据，提高分发引擎的响应速度。本文利用数据库 DML 触发器技术，针对基础数据表做出监控，将变更信息记录到单独的索引模块，数据同

表 2 各功能模块说明

序号	模块	功能描述
1	触发器模块	触发器模块包括函件、文件、附件、关系和合同等触发器，利用数据库 DML 触发器技术，对于这些内容媒体的增加、删除、修改等信息会实时地记录到变更信息记录模块，形成变更信息索引
2	变更索引模块	记录内容媒体信息的变更，将海量数据的查询操作转换为对此模块的查询，大大降低内容数据抓取复杂度
3	分发规则配置模块	对分发业务进行高度抽象，公司之间的分发业务均可由此表配置，一条记录对应一条分发规则
4	分发规则计算模块	根据内容媒体的渠道号信息，结合分发规则配置模块，自动计算出每天内容媒体的分发类型

步程序直接根据索引模块来实现数据同步，数据量实现大大降级，提高抓取效率。内容抓取模块交互示意图如图 5 所示。

3.2.1 信息记录变更

针对上述内容媒体，设计的内容分发索引库结构见表 3 和表 4。文档人员对 AED_ECM 做出

的改动（增、删、改），会通过 DML 触发器（见表 5），自动记录到 AED_INDEX 库，并通过 FLAG 字段进行同步标识。

3.2.2 抓取过程

内容分发引擎按照如下算法进行内容抓取，如图 6 所示。利用此机制，由原来的对 AED_ECM

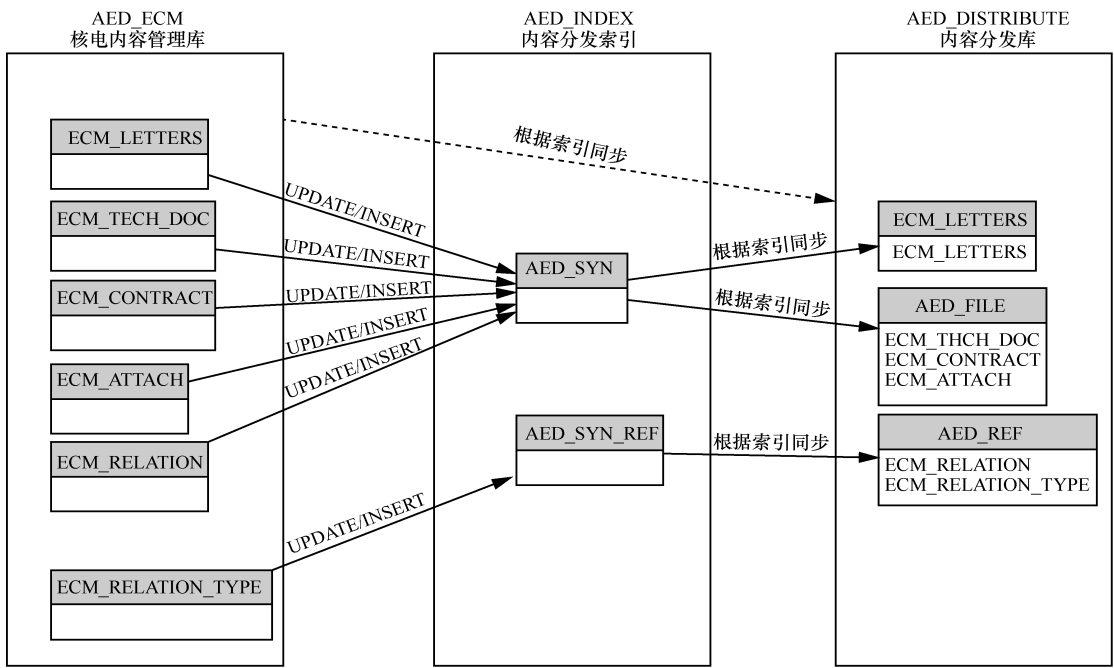


图 5 内容抓取模块交互示意图

表 3 AED_SYN 表结构

表字段	类型	是否主键	是否可空	默认值	备注
ID	int	是	否		内容索引表主键 ID
CATEGORY	nvarchar(10)	否	否		内容媒体类型：letter/file/att/cont
RID	nvarchar(50)	否	否		内容唯一标识符 R_OBJECT_ID
RTYPE	nvarchar(50)	否	否		函件类型，用于区别传递单函件和普通函件
FLAG	int	否	否	0	同步标识，0 表示未同步，1 表示已同步



表 4 AED_SYN_REF 表结构

表字段	类型	是否主键	是否可空	默认值	备注
ID	int	是	否		内容关系索引表主键 ID
RID	nvarchar(50)	否	否		内容关系唯一标识符 R_OBJECT_ID
FLAG	int	否	否	0	同步标识, 0 表示未同步, 1 表示已同步

表 5 AED_ECM DML 触发器

DML Trigger	功能描述
Tr_Syn_ECM_LETTERS	抓取 ECM_LETTERS 表的变更, 在 AED_SYN 中记录索引
Tr_Syn_ECM_TECH_DOC	抓取 ECM_TECH_DOC 表的变更, 在 AED_SYN 中记录索引
Tr_Syn_ECM_ATTACH	抓取 ECM_ATTACH 表的变更, 在 AED_SYN 中记录索引
Tr_Syn_ECM_CONTRACT	抓取 ECM_CONTRACT 表的变更, 在 AED_SYN 中记录索引
Tr_Syn_ECM_RELATION	抓取 ECM_RELATION 表的变更, 在 AED_SYN_REF 中记录索引

中的表进行全表扫描, 转变成对 AED_INDEX 索引表进行操作。索引表数据量很小, 大大降低了获取分发数据的时间。

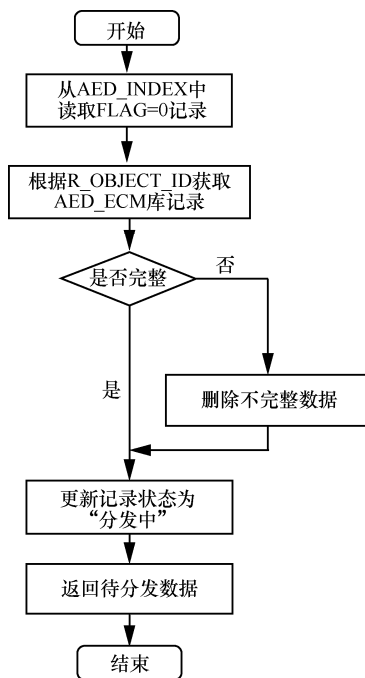


图 6 内容抓取流程

3.3 引擎设计

对于内容媒体的接收和发送, 由于发送方/接收方不同, 启动的分发类型也不相同。同一公司在和不同的公司交互资料时, 使用的三字码/四字

码代字也各不相同。本文利用分发规则库来存储这些分发规则, 使得分发规则的管理得到高度统一, 而且可以根据业务进行即时配置, 灵活应对业务的变更。

3.3.1 分发规则库

AED 平台中将分发规则设置为每个公司的基本属性, 利用公司表 AED_COMPANY 和分发类型表 AED_DISTRTYPE 来存储每个公司的分发规则, 表结构设计见表 6 和表 7。

上述表设计可支持所有分发业务, 同时也能应对业务变更。根据 AED 分发业务, AED 内容分发引擎分发规则库配置见表 8 和表 9。

3.3.2 分发规则计算算法

内容媒体依据不同类别具有不同的编码格式, 但均具有项目代码、发送方、接收方、流水号等基本信息, 以在建项目的外部信函编码为例, 其编码格式和主要属性构成如图 7 和表 10 所示。

结合分发规则库, 按照图 8 所示的流程来判断内容媒体的分发类型 (以函件 “BL-AEQ-707132-ANS” 为例)。首先判断内容媒体的分发模式, 发送方或接收方有一个为外部公司的, 即判定为外部发送/外部接收, 其他均为内部发送/内部接收; 然后查询分发规则库, 判断启动分发类型 1

表 6 AED_COMPANY 表结构设计

表字段	类型	字段意义	可空	备注
Id	int	主键	否	
EpCode	nvarchar(50)	公司代码	否	
EpName	nvarchar(50)	公司名称	是	
NeedSZM	nvarchar(4000)	三字码/通信代码	是	可使用的三字码, 逗号分隔多个三字码
IsNB	bit	是否为内部公司	是	标示是系统内的公司还是第三方公司, 1 是内部公司, 0 是外部公司
NBSend	int	内部发送	是	系统内公司间互发, 发送时对应的分发类型、AED_DISTRTYPE 主键值
NBReceive	int	内部接收	是	系统内公司间互发, 接收时对应的分发类型、AED_DISTRTYPE 主键值
WBSend	int	外部发送	是	系统外, 发送时对应的分发类型、AED_DISTRTYPE 主键值
WBReceive	int	外部接收	是	系统外, 接收时对应的分发类型、AED_DISTRTYPE 主键值
IsFFInfo	bit	分发信息	是	分发信息是否参与判断。如果参与判断, 则按照分发信息(主办、协办、阅知)启动自动分发; 否则, 按 WBReceive 中配置值来启动一次分发
SendCode	nvarchar(50)	默认发送方	是	
Remark	nvarchar(4000)	备注	是	

表 7 AED_DISTRTYPE 表结构设计

表字段	类型	字段意义	可空	备注
FFTypeId	int	主键	否	分发类型表主键
OneFFType	nvarchar(50)	一次分发类型	是	设总、CAC、文档总部/现场、自动分发
FFName	nvarchar(50)	分发名称	是	分发名称
ToCode	nvarchar(50)	接收方名称	是	展示名称

表 8 AED_COMPANY 表配置

字段规则	设计院规则	工程公司总部规则	陆丰公司	工程公司现场规则
Id	1	2	3	4
EpCode	CNPDC	CNPEC	LFNP	CNPEC
EpName	设计公司	工程公司	陆丰公司	工程公司
NeedSZM	AEQ,YAEQ,AES	ANE,YANE,GENS	LFNP	ANS,GBNS,YANS
IsNB	1	1	1	1
NBSend	1	1	1	1
NBReceive	2	3	5	4
WBSend	1	1	1	1
WBReceive	2	3	5	4
IsFFInfo	0	1	0	1
SendCode	AEQ		LFNP	
Remark		总部		现场



表 9 AED_DISTRTYPE 表配置

字段类型	分发类型 1	分发类型 2	分发类型 3	分发类型 4	分发类型 5
FFTypeId	1	2	3	4	5
OneFFType	自动分发	设总分发	工程总部文档分发	工程现场文档分发	业主文档分发
FFName	自动分发	设总分发	CAC 分发	CAC 分发	CAC 分发
ToCode		设总分发	CAC 分发	CAC 分发	CAC 分发

和 4，即启动自动分发和工程公司现场文档分发；最后根据分发信息 IsFFInfo 是否生效，判断只启动自动分发。

表 10 主要属性构成

函件编码	...	发送方	接收方	...	主办	协办	阅知
NN-XXX-	...	XXX	YYY	...	张三	李四	王五
NNNNNN-YYYY							

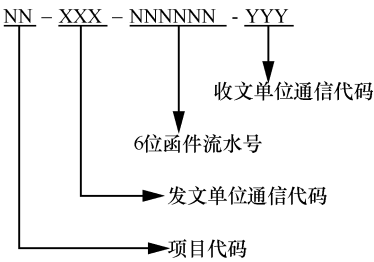


图 7 编码格式

具体的分发规则计算算法用伪代码如下。

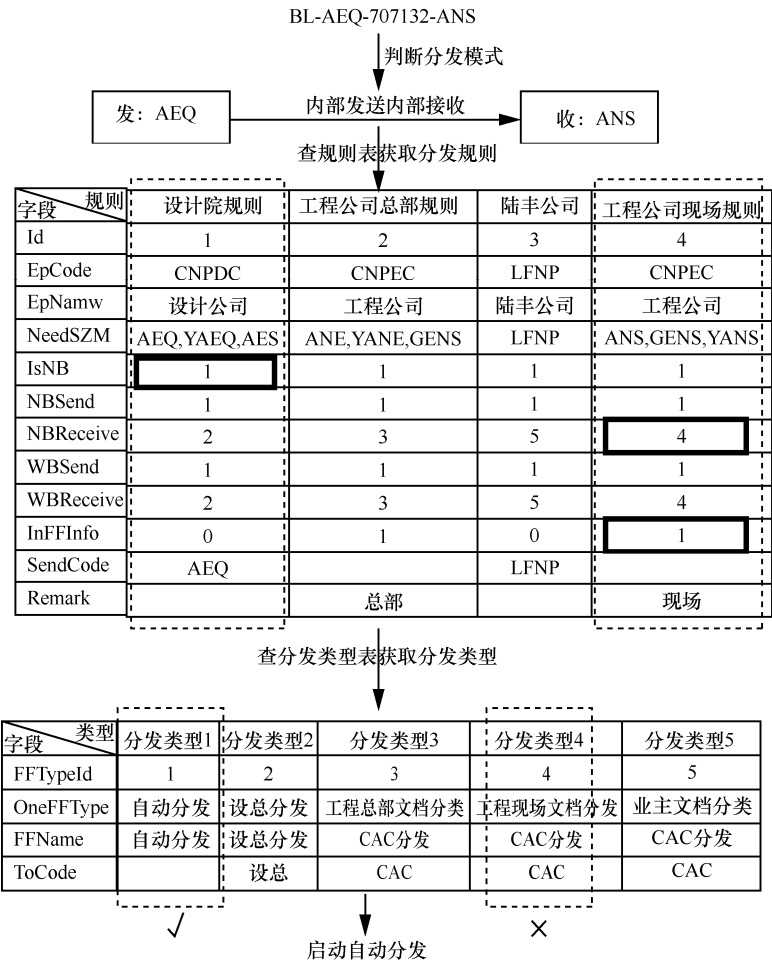


图 8 内容媒体的分发类型的判断流程

根据发送方、接收方、主办、协办、阅知，
计算内容媒体分发类型

COMPUTE-DISTRTYPE(F, T, Z, X, Y)

输入： F 为发送方， T 为接收方， Z 为主办，
 X 为协办， Y 为阅知

输出：分发类型

SZMs $\leftarrow$ split(NeedSZM)

RArray(type, ffinfo) $\leftarrow$ (0,0)

if F in SZMs //如果发送方在三字码列表中

then

从 AED_COMPANY 中找出 F 对应的
(NBReceive, 0) 值 N_VF ;

if T in SZMs //如果接收方在三字码列表中，
则为内部发送—内部接收

then

从 AED_COMPANY 中找出 T 对应的
(NBSend, 0) 值 N_VT ;

insert N_VT, N_VF into RArray;

end

else //如果接收方不在三字码列表中，则为外
部发送—外部接收

从 AED_COMPANY 中找出 F 对应的
WBSend 值 W_VF ;

Insert W_VF into RArray;

end

end

else //如果发送方不在三字码列表中

then

从 AED_COMPANY 中找出 F 对应的
(WBReceive, IsFFInfo) 值 W_VF ;

Insert W_VF into RArray;

end

for a in RArray//如果分发信息不为空且接收
方 T 对应的分发信息生效，则只启动自动分发

then

if $Z \neq "& & X \neq "& & Y \neq "$ and IsFFInfo=1

将 RArray 的 type 全部置为 1;

end

return RArray;

从上述算法伪代码可知，分发规则计算算法的时间复杂度和空间复杂度均为 $O(N)$ ，其中 N 为分发体系中分发对象的个数。在生产环境中，单条内容媒体的分发规则平均计算时间仅为 0.003 s，这种高效的线性计算算法，大大保证了启动一次分发的效率。

4 内容分发引擎的应用

本文设计的内容分发引擎已经在中广核工程有限公司的核电工程文档管理与分发平台(AED)中投产，内容分发引擎承担了前期的数据抓取和启动一次分发的任务。目前平均每天由文档人员手工录入和各业务系统归档进入 AED 平台的内容媒体数量超过 1 000 份，从投产至今已分发了各类内容媒体 1 095 000 份，总的分发次数达到 23 867 008 次，平均日分发次数达到 21 796.4 次。表 11 给出了相比实施前的分发平台的相关 KPI 对比分析。

本方案首次引入分发规则库和配置化思想，分发引擎具有极高的可扩展性和通用性。若公司因业务变更而导致分发规则的变更，只需修改表 12 和表 13 中的对应的对象和规则即可，平

表 11 实施效果 KPI 对比

KPI	本方案实施前	本方案实施后
内容抓取即时性	3.5 h/1 000 份	2.6 min/1 000 份
平台吞吐量(日分发次数)	7 265.5 次/天	21 796.4 次/天
平台可扩展性	扩展性极差(修改代码)	高可扩展性(增加配置)
平台通用性	无法通用	高通用性
服务部署	分发引擎硬编码，随 Web 站点部署，无法独立部署	分发引擎服务可独立部署



表 12 分发规则库—对象配置

字段	设计院规则	工程公司总部规则	陆丰公司	工程公司现场规则	WECAN公司
Id	1	2	3	4	WECAN
EpCode	CNPDC	CNPEC	LFNP	CNPEC	WECAN公司
EpNamw	设计公司	工程公司	陆丰公司	工程公司	AWQ,WENS
NeedsZM	AEQ,YAEQ,AES	ANE,YANE,GENS	LFNP	ANS,GENS,YANS	1
IsNB	1	1	1	1	1
NBSend	1	1	1	1	6
NBReceive	2	3	5	4	1
WBSend	1	1	1	1	6
WBReceive	2	3	5	4	0
InFFInfo	0	1	0	1	AW0
SendCode	AEQ		LFNP		
Remark		总部		现场	

表 13 分发规则库—分发规则配置

字段	分发类型1	分发类型2	分发类型3	分发类型4	分发类型5	分发类型
FFTypeId	1	2	3	4	5	6
OneFFType	自动分发	设总分发	工程总部文档分类	工程现场文档分发	业主文档分类	WECAN发布
FFName	自动分发	设总分发	CAC分发	CAC分发	CAC分发	设总分布
ToCode		设总	CAC	CAC	CAC	

台代码层面无需任何变更。另外，对于新公司加入分发体系，只需根据业务逻辑在分发规则库中进行配置，无需修改分发引擎。例如，WECAN公司作为设计单位也参与分发体系，只需在分发规则库中进行表 12 和表 13 的配置，WECAN 公司便可参与内外部公司的内容分发。

5 结束语

本文设计了一种可配置化的核电企业内容分发引擎，该分发引擎能够高效地抓取待分发内容媒体，支持业务逻辑配置化，减少了由分发业务变化而导致的引擎频繁变更，具有高通用性、高实时性和高扩展性。在生产环境中，内容分发引擎表现出极高的吞吐量，运行平稳、效果良好。不仅限于核电企业内容管理工作，其设计思想对于其他领域的核心内容分发工作也提供了指导与参考作用。

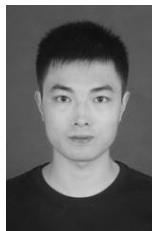
参考文献：

- [1] 武钺. 全电子化策略下的核电工程文档管理信息系统建设[J]. 中国档案, 2012(6): 64-65.
WU Y. Construction of nuclear power engineering document management information system under full electronic strategy[J]. China Archives, 2012(6): 64-65.
- [2] 汪硕. 一个文件分发系统的设计与实现[D]. 北京: 北京大学, 2013.

WANG S. Design and implementation of a file distribution system[D]. Beijing: Peking University, 2013.

- [3] 中广核工程有限公司. 文件管理规定[R]. 2016.
China Nuclear Power Engineering Co., Ltd. Document management regulations[R]. 2016.
- [4] 杨明川. 内容分发网络关键技术分析[J]. 电信科学, 2005, 21(8): 13-17.
YANG M C. Key technology researches of content distribution network[J]. Telecommunications Science, 2005, 21(8): 13-17.

[作者简介]



汪鹏（1988—），男，深圳中广核工程设计有限公司核电设计平台建设主管工程师，主要研究方向为核电企业内容管理平台开发、工作流开发、项目管理。



涂红兵（1979—），男，深圳中广核工程设计有限公司核电设计平台开发与管理高级工程师、信息开发室主任工程师，主要研究方向为核电生产平台架构、云计算、工作流。

王云福（1973—），男，深圳中广核工程设计有限公司核电设计平台开发与管理高级工程师、信息开发室主任，主要研究方向为信息系统审计、软件研发项目管理。