

BỘ THÔNG TIN VÀ TRUYỀN THÔNG
VIỆN CÔNG NGHIỆP PHẦN MỀM VÀ NỘI DUNG SỐ VIỆT NAM

BÁO CÁO TÓM TẮT
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ

NGHIÊN CỨU XÂY DỰNG TIÊU CHUẨN VỀ ĐỊNH DANH NỘI DUNG SỐ
DÙNG TRONG LƯU TRỮ VÀ TRAO ĐỔI DỮ LIỆU

TRÊN MÔI TRƯỜNG MẠNG

Mã số: 31-15-KHKT-TC

Hà Nội - 2015

Mục lục

1 GIỚI THIỆU	3
1.1 Tên đề tài.....	3
1.2 Mã số.....	3
1.3 Mục tiêu.....	3
1.4 Nội dung.....	3
1.5 Kết quả.....	3
2 SỰ CẦN THIẾT	3
3 NGHIÊN CỨU MỘT SỐ TIÊU CHUẨN QUỐC TẾ VÀ QUỐC GIA VỀ ĐỊNH DANH NỘI DUNG SỐ	6
3.1 Trong nước.....	6
3.2 Ngoài nước.....	7
3.3 Kết luận.....	8
4 MỘT SỐ PHƯƠNG PHÁP ĐÁNH MÃ ĐỊNH DANH	8
4.1 Phương pháp đánh mã định danh Unixtime.....	8
4.2 Phương pháp đánh mã định danh UUID.....	9
4.3 Kết luận.....	9
5 NGHIÊN CỨU, ĐỀ XUẤT TIÊU CHUẨN ĐỊNH DANH CHO VĂN BẢN HÀNH CHÍNH CÔNG ĐIỆN TỬ	10
5.1 Mục đích xây dựng tiêu chuẩn.....	10
5.2 Mục tiêu xây dựng tiêu chuẩn.....	10
5.3 Phương pháp luận xây dựng tiêu chuẩn.....	10
5.4 Cấu trúc dự thảo tiêu chuẩn TCVN xxxx:201x.....	11
6 KẾT LUẬN VÀ KIẾN NGHỊ	12

1 GIỚI THIỆU

1.1 Tên đề tài

Nghiên cứu xây dựng tiêu chuẩn về định danh nội dung số dùng trong lưu trữ và trao đổi dữ liệu trên môi trường mạng

1.2 Mã số

31-15-KHKT-TC

1.3 Mục tiêu

Giúp cho việc trao đổi, tìm kiếm tài liệu một cách thuận tiện dễ dàng.

1.4 Nội dung

Nghiên cứu một số tiêu chuẩn quốc tế và quốc gia về định danh nội dung số

Nghiên cứu phương pháp đánh mã số định danh Unixtime

Nghiên cứu phương pháp đánh mã số định danh UUID

Nghiên cứu đề xuất tiêu chuẩn định danh cho văn bản hành chính công điện tử

Xây dựng dự thảo tiêu chuẩn định danh cho văn bản hành chính công điện tử

1.5 Kết quả

Báo cáo kết quả đề tài, thuyết minh và dự thảo tiêu chuẩn về định danh cho văn bản hành chính công điện tử dùng trong lưu trữ và trao đổi dữ liệu trên môi trường mạng.

2 SỰ CẦN THIẾT

Tài liệu là một nguồn thông tin không thể thiếu nhằm phục vụ hoạt động của xã hội và con người. Thông tin được gửi và nhận bằng nhiều phương tiện khác nhau, trong đó thể hiện dưới dạng văn bản là một phương tiện chủ yếu và phổ biến nhất. Văn bản vừa là công cụ, vừa là phương tiện cơ bản và chủ yếu để các cơ quan, tổ chức trao đổi thông tin, phục vụ hoạt động của các cơ quan. Thực tiễn cho thấy, tất cả các cơ quan, tổ chức, cá nhân... không thể không sử dụng văn bản trong các hoạt động giao dịch, điều hành tổ chức với một khối lượng lớn. Trong công tác quản lý nhà nước, văn bản là công cụ giao tiếp thành văn, quan trọng giữa cơ quan nhà nước với các cá nhân, tổ chức khác, là phương tiện quan trọng để ghi lại và truyền đạt các quyết định quản lý,

là hình thức để cụ thể hóa pháp luật, là phương tiện để điều chỉnh những quan hệ xã hội thuộc phạm vi quản lý nhà nước.

Trong các cơ quan, tổ chức văn bản hành chính chiếm số lượng lớn trong hệ thống văn bản của cơ quan, đơn vị, nó vừa có ý nghĩa pháp lý vừa có ý nghĩa thực tiễn trong quản lý nhà nước, quản lý kinh tế.

Sự bùng nổ của CNTT-TT song hành với sự gia tăng chóng mặt của thông tin, nội dung số trong cơ quan, tổ chức, các quy trình nghiệp vụ hành chính cũng dần được tin học hóa. Nhu cầu chia sẻ dữ liệu, trao đổi thông tin và tương tác giữa các hệ thống thông tin và giữa các cơ quan, tổ chức ngày càng tăng để tránh lãng phí tài nguyên, tiết kiệm thời gian. Tuy nhiên, việc trao đổi thông tin và hoạt động thông qua môi trường điện tử của các đơn vị còn hạn chế mà lý do cốt lõi là chưa có hệ thống các văn bản qui định đầy đủ về dữ liệu, quản lý, điều hành, trao đổi VBDT qua mạng đảm bảo dữ liệu được liên thông thông suốt giữa các hệ thống.

Hiện nay, phần lớn các văn bản hành chính trao đổi giữa các cơ quan, tổ chức với nhau vẫn bằng hình thức giấy tờ, gây nhiều lãng phí, chưa khai thác tận dụng được những điều kiện hiện có về hạ tầng CNTT, chưa tạo nên được môi trường làm việc điện tử hiện đại, minh bạch, giảm thủ tục giấy tờ, tiết kiệm chi phí, thời gian.

Vì những hạn chế, tồn tại trên, dẫn đến việc ứng dụng CNTT trong hoạt động trao đổi văn bản trên môi trường điện tử mới ở bề nổi, chưa thực sự đi vào chiều sâu, tạo ra điểm nghẽn trong nâng cao hiệu quả ứng dụng CNTT cũng như hiệu quả quản lý, hoạt động của các đơn vị.

Do văn bản là công cụ hiệu quả được sử dụng trong mối quan hệ tương tác giữa các cơ quan, tổ chức với nhau cùng với ý nghĩa và tầm quan trọng của việc sử dụng VBDT trong hoạt động cơ quan, tổ chức, do đó đòi hỏi cần phải có các giải pháp thích hợp để giúp các cơ quan quản lý các nguồn thông tin, văn bản một cách khoa học, phục vụ việc chỉ đạo, điều hành, hỗ trợ ra quyết định, truy cứu trách nhiệm, lưu trữ.... ***Do vậy, việc xây dựng các tiêu chuẩn liên quan đến văn bản điện tử là hết sức quan***

trọng và ngày càng cấp thiết. Việc chậm trễ ban hành các quy định liên quan đến văn bản điện tử sẽ làm cho các ứng dụng phát triển về sau này, đặc biệt là các phần mềm phục vụ cho các dịch vụ cung cấp dữ liệu không sử dụng được các dữ liệu trước đây, làm cản trở việc trao đổi và cung cấp thông tin – gây ra tình trạng cát cứ thông tin. Một trong số các tiêu chuẩn liên quan đến văn bản điện tử phải kể đến là tiêu chuẩn về định danh văn bản. Trong phạm vi nghiên cứu của đề tài sẽ đưa ra qui định về việc gán định danh cho văn bản hành chính công điện tử. ***Mục tiêu chính của việc xây dựng tiêu chuẩn định danh cho văn bản hành chính công điện tử là để đảm bảo tính duy nhất của văn bản trong hệ thống văn bản hành chính công điện tử của các cơ quan, tổ chức (tức là đảm bảo rằng định danh cấp cho hai đối tượng khác nhau là khác nhau) dùng trong lưu trữ, quản lý văn bản trên môi trường mạng, đồng thời cũng để xác định được đối tượng tham gia quá trình gửi nhận văn bản giữa các hệ thống trên môi trường mạng.***

Như chúng ta thấy, đối tượng liên quan trong hoạt động trao đổi văn bản là đối tượng tác nhân (con người/tổ chức/hệ thống) và đối tượng văn bản. Các đối tượng này cần được gán mã để quản lý, xác thực. Việc đánh mã các đối tượng tham gia trao đổi văn bản điện tử xuất phát từ nhu cầu thực tế về kết nối các hệ thống QLVBĐH phục vụ việc trao đổi văn bản điện tử của tất cả các cơ quan trong hệ thống chính trị của Việt Nam. Việc đánh mã là để xác định duy nhất các đối tượng tham gia trao đổi văn bản điện tử. ***Do vậy trong Đề tài này cần lựa chọn phương pháp gán mã định danh cho văn bản hành chính công điện tử mà khi phân giải chuỗi ký tự định danh có thể cho biết (khi phục) được thông tin về nơi phát hành văn bản.***

Hệ thống định danh được thiết kế để ứng dụng trên Internet. Một định danh duy nhất được gán vĩnh viễn cho một đối tượng để đảm bảo một liên kết mạng liên tục đến thông tin về đối tượng đó, kể cả trường hợp thông tin về đối tượng có thể thay đổi theo thời gian thì định danh duy nhất của đối tượng là không thay đổi.

Ở một số hệ thống lưu trữ dữ liệu lớn như Googledoc, hệ thống quản trị nội dung số Alfresco,

các tài liệu cũng được quản lý theo định danh ID của tài liệu. Để lưu trữ và quản lý chia sẻ các tài liệu, mỗi tài liệu được tạo ra được gán một định danh duy nhất. Định danh này của tài liệu không thay đổi bất kể có những thay đổi được thực hiện trên siêu dữ liệu và/hoặc nội dung của tập tin.

Đối với hệ thống văn bản hành chính hiện nay, với đặc điểm của văn bản hành chính có các thông tin cơ bản bao gồm: Số ký hiệu văn bản, cơ quan ban hành văn bản, thời gian ban hành văn bản. Sự tổ hợp của ba yếu tố cơ bản này đủ để nhận diện văn bản hành chính trong hệ thống các văn bản hành chính của cơ quan, tổ chức. Tuy nhiên để quản lý được hệ thống các văn bản này cần phải có đầy đủ tối thiểu ba thông tin cơ bản nêu trên. Khi mà một mã định danh được gán cho văn bản hành chính thì việc quản lý đối tượng văn bản hành chính sẽ hoàn toàn dựa theo mã định danh của văn bản.

3 NGHIÊN CỨU MỘT SỐ TIÊU CHUẨN QUỐC TẾ VÀ QUỐC GIA VỀ ĐỊNH DANH NỘI DUNG SỐ

3.1 Trong nước

Hiện nay ở Việt Nam, trong xu thế phát triển toàn cầu và hội nhập, các nghiên cứu và kết quả đa phần là Việt hóa các chuẩn quốc tế hiện hành. Trong lĩnh vực thông tin và truyền thông nói chung và Thông tin và tư liệu nói riêng, Việt Nam đã ban hành một số tiêu chuẩn về mã số, định danh cho sản phẩm nội dung số như sau:

- TCVN 6380:2007 – Tiêu Chuẩn Việt Nam – Thông tin và tư liệu. Mã số tiêu chuẩn Quốc tế cho sách (ban hành ngày 31-12-2007 theo quyết định số 3269/QĐ-BKHHCN của Bộ Khoa học và Công nghệ)
- TCVN 6381:2007 - Thông tin và tư liệu. Mã số tiêu chuẩn quốc tế cho xuất bản phẩm nhiều kỳ (ban hành ngày 31-12-2007 theo quyết định số 3247/QĐ-BKHHCN của Bộ Khoa học và Công nghệ)

Ngoài một số tiêu chuẩn về mã số cho một số đối tượng thuộc sản phẩm nội dung số nêu trên, hiện Việt Nam chưa có một tiêu chuẩn hay quy định nào về cấp (gán) mã định danh cho đối tượng văn bản nói chung và văn bản điện tử nói riêng.

Gần đây Cục Văn thư và lưu trữ nhà nước – Bộ Nội vụ có ra văn bản số 822/HD-VTLTNN ngày 26/8/2015 về Hướng dẫn quản lý văn bản đi, văn bản đến và lập hồ sơ trong môi trường mạng. Theo đó văn bản có hướng dẫn về gán Mã hồ sơ lưu trữ văn bản đi, văn bản đến như sau:

Dùng 14 ký tự số để gán mã hồ sơ lưu trữ văn bản đi, văn bản đến. Các ký tự số từ trái qua phải được phân bổ như sau:

- Các ký tự số thứ 1, thứ 2, và thứ 3 là mã số của cơ quan, tổ chức.
- Các ký tự số thứ 4, thứ 5, thứ 6 và thứ 7 là mã số thể hiện năm bắt đầu chuyển văn bản sang chế độ lưu trữ, năm này được xác định là năm công việc kết thúc.
- Các ký tự số thứ 8, thứ 9, thứ 10 là mã số được chia theo nhóm tài liệu quản lý trong mỗi cơ quan theo quy định tại Thông tư 09/2011/TT-BNV ngày 03/6/2011 của Bộ Nội vụ quy định về thời hạn bảo quản hồ sơ, tài liệu hình thành phổ biến trong hoạt động của các cơ quan, tổ chức.
- Các ký tự số thứ 11, thứ 12, thứ 13 là mã số đại diện cho nhóm nhỏ tài liệu theo từng lĩnh vực quản lý chuyên môn trong mỗi cơ quan, tổ chức
- Ký tự số thứ 14 là số kiểm tra

Tuy nhiên hiện chưa có qui định hay hướng dẫn cụ thể nào về gán mã định danh cho đối tượng văn bản nói chung và văn bản hành chính công điện tử nói riêng.

3.2 Ngoài nước

Cho đến thời điểm hiện tại, tổ chức tiêu chuẩn quốc tế ISO đã xây dựng và ban hành khá nhiều tiêu chuẩn về gán mã số, định danh nói chung và mã số, định danh cho một

số sản phẩm nội dung số nói riêng. Danh sách các tiêu chuẩn về mã số, định danh cho một số sản phẩm nội dung số bao gồm:

- ISMN – mã số tiêu chuẩn quốc tế cho âm nhạc
- ISAN - Mã số tiêu chuẩn quốc tế cho tác phẩm nghe nhìn
- ISBN - Mã số tiêu chuẩn quốc tế cho sách (ISO 2108)
- ISSN - Mã số tiêu chuẩn quốc tế cho xuất bản phẩm nhiều kỳ (ISO 3297)
- ISO 26324 - Information and documentation - Digital object identifier system (ISO 26324)

Tuy nhiên trong hệ thống các tiêu chuẩn của nước ngoài hiện chưa có một tiêu chuẩn cụ thể nào về định danh cho đối tượng văn bản điện tử nói chung và văn bản hành chính công điện tử nói riêng.

3.3 Kết luận

Qua nghiên cứu một số tiêu chuẩn quốc tế và quốc gia về gán mã số, định danh cho một số sản phẩm nội dung số cho thấy:

- (1) Hầu hết các tiêu chuẩn đều có giới hạn nhất định về phạm vi loại tài liệu, ví dụ như cho sách (ISO 2108), cho xuất bản phẩm nhiều kỳ (ISO 3297)..., Và hầu hết các mã số gán cho các đối tượng được tạo ra theo nguyên tắc có sự kết hợp của yếu tố về Nguồn gốc của đối tượng (tức là thông tin về Nhà xuất bản hay nhà phát hành ra ấn phẩm..). Các tiêu chuẩn này phần lớn đã được xây dựng và ban hành thành Tiêu chuẩn quốc gia Việt Nam.
- (2) Tiêu chuẩn quốc tế ISO 26324:2012 có thể được xem là bộ tiêu chuẩn phù hợp cho việc định danh các đối tượng không bị giới hạn bởi loại dữ liệu, hình thức dữ liệu, phù hợp với việc phát triển và trao đổi thông tin. Tiêu chuẩn này cũng chưa được xây dựng và ban hành thành Tiêu chuẩn Quốc gia TCVN

4 MỘT SỐ PHƯƠNG PHÁP ĐÁNH MÃ ĐỊNH DANH

4.1 Phương pháp đánh mã định danh Unixtime

Unix time: là một hệ thống để mô tả một khoảnh khắc thời gian trong khoảng thời gian, được định nghĩa là số giây đã trôi qua kể từ 00:00:00 (Coordinated Universal Time - UTC).

Phương thức mã hóa Unix time: Có 2 lần mã hóa trong phương pháp Unixtime. 2 lần được tách ra riêng biệt: LỚp đầu tiên giải mã một điểm trong thời gian như một số thực vô hướng, và lần thứ hai mã hóa số đó như một chuỗi các bit hoặc chữ số thập phân

Phương pháp đánh mã số định danh Unixtime: Phương pháp đánh mã số định danh Unixtime là phương pháp gán chuỗi chữ số thập phân được mã hóa từ một điểm thời gian tính theo giây cho một đối tượng làm mã số định danh cho đối tượng đó.

4.2 Phương pháp đánh mã định danh UUID

UUID là viết tắt của Universally Unique Identifier. Nó là một chuẩn định danh thường được dùng trong các kiến trúc phần mềm và được chuẩn hóa bởi Open Software Foundation (OSF)

Cách tạo ra một dãy số UUID:

- Kết hợp giữa thời gian với địa chỉ MAC
- Sử dụng pseudo-random
- Sử dụng mật mã và các chuỗi ký tự văn bản được cung cấp.

Kết quả các UUID được sinh ra như vậy sẽ mang tính chất là duy nhất

(Tạo ra 1 tỷ UUID mỗi giây trong vòng 100 năm tới thì xác suất bị trùng sẽ khoảng 50%)

Phương pháp đánh mã số định danh UUID: Là phương pháp gán chuỗi ký tự được tạo ra từ các yếu tố đầu vào cho một đối tượng làm mã số định danh cho đối tượng đó

4.3 Kết luận

Tại cùng một thời điểm mà hai văn bản cùng được tạo ra và được gán định danh theo phương pháp Unixtime thì hai văn bản này sẽ có cùng mã số định danh. Để đảm bảo việc gán mã số định danh cho hai đối tượng này là khác nhau cần kết hợp thêm các

yếu tố khác để đảm bảo phân biệt được các văn bản trong hệ thống văn bản của các cơ quan, tổ chức chẳng hạn kết hợp với trường “Mã số cơ quan, tổ chức ban hành văn bản” làm thành phần tiền tố cho chuỗi ký tự được gán làm định danh cho văn bản. Đối với phương pháp UUID, với bất kỳ đầu vào khác nhau thì sẽ cho chuỗi đầu ra khác nhau và đảm bảo tính duy nhất. Tuy nhiên với chuỗi ký tự được tạo ra theo phương pháp UUID thì khó có thể thực hiện việc phân giải ngược ra các yếu tố dữ liệu đầu vào. Với mục tiêu của đề tài là đưa ra qui định về việc gán định danh cho văn bản hành chính công điện tử để đảm bảo tính duy nhất của văn bản trong hệ thống văn bản hành chính công điện tử của các cơ quan, tổ chức. Đề tài không giải quyết bài toán ngược lại từ định danh của văn bản để truy xuất ngược lại các thông tin tạo nên định danh duy nhất cho văn bản. Do đó, đề tài lựa chọn phương pháp UUID là một trong số phương pháp để tạo mã định danh cho văn bản hành chính công điện tử

5 NGHIÊN CỨU, ĐỀ XUẤT TIÊU CHUẨN ĐỊNH DANH CHO VĂN BẢN HÀNH CHÍNH CÔNG ĐIỆN TỬ

5.1 Mục đích xây dựng tiêu chuẩn

Xây dựng tiêu chuẩn nhằm bổ sung thêm tiêu chuẩn vào hệ thống tiêu chuẩn về lĩnh vực Thông tin và tư liệu hiện còn đang thiếu nhiều ở Việt Nam để khuyến nghị áp dụng tại Việt Nam.

5.2 Mục tiêu xây dựng tiêu chuẩn

Nghiên cứu, đưa ra qui định về việc gán định danh cho văn bản hành chính công điện tử để đảm bảo tính duy nhất của văn bản trong hệ thống văn bản hành chính công điện tử của các cơ quan, tổ chức (tức là đảm bảo rằng định danh cấp cho hai đối tượng khác nhau là khác nhau) dùng trong lưu trữ, quản lý và trao đổi văn bản trên môi trường mạng.

Với mục tiêu để lưu trữ, quản lý và trao đổi văn bản hành chính công điện tử trên môi trường mạng, do vậy, định danh gán cho văn bản hành chính công điện tử không cần thiết phải dễ đọc, dễ hiểu.

5.3 Phương pháp luận xây dựng tiêu chuẩn

Trên cơ sở nghiên cứu, phân tích một số tiêu chuẩn quốc tế và quốc gia về gán mã số, định danh cho đối tượng nội dung số; một số phương pháp đánh mã số định danh, nhóm dự thảo tiêu chuẩn đã chọn phương pháp UUID để tạo chuỗi ký tự gán làm mã định danh cho văn bản hành chính công điện tử, đồng thời lựa chọn tài liệu ISO 26324:2012 làm tài liệu viện dẫn chính để biên soạn dự thảo TCVN, trong đó, thành phần hậu tố của chuỗi ký tự được gán làm định danh cho văn bản hành chính công điện tử có thể được tạo ra bằng phương pháp unixtime, phương pháp uuid, hoặc các phương pháp sinh khác.

5.4 Cấu trúc dự thảo tiêu chuẩn TCVN xxxx:201x

Tiêu chuẩn này gồm 09 điều và 03 phụ lục cụ thể như sau:

- 1 Phạm vi áp dụng
 - 2 Tài liệu viện dẫn
 - 3 Thuật ngữ và định nghĩa
 - 4 Tiêu chí gán định danh
 - 5 Định danh DOI
 - 6 Gán định danh cho văn bản hành chính công điện tử theo phương pháp UUID
 - 7 Gán định danh DOI cho văn bản hành chính công điện tử
 - 8 Phân giải định danh DOI0
 9. Siêu dữ liệu DOI
 - 10 Quản lý hệ thống DOI
- Phụ lục I: Mối quan hệ giữa hệ thống DOI và các sơ đồ định danh khác
- Phụ lục II: Đặc tính kỹ thuật của siêu dữ liệu DOI cho tài liệu
- Phụ lục III: Quản lý hệ thống DOI

6 KẾT LUẬN VÀ KIẾN NGHỊ

Mục tiêu của tiêu chuẩn TCVN xxxx:201X này thiết lập các qui định kỹ thuật về định danh cho văn bản hành chính công điện tử như là hệ thống phân định đơn nhất cho mỗi văn bản hành chính công điện tử trong hệ thống các văn bản hành chính công điện tử của các cơ quan, tổ chức.

Tiêu chuẩn này được xây dựng dựa trên tiêu chuẩn ISO 26324:2012 là bộ tiêu chuẩn cho việc định danh các đối tượng không bị giới hạn bởi loại dữ liệu, hình thức dữ liệu, phù hợp với việc phát triển và trao đổi thông tin. Tiêu chuẩn này được ban hành sẽ giúp cho các cơ quan, tổ chức có thể dựa vào đó để thực hiện việc đánh mã định danh cho các văn bản hành chính công điện tử phục vụ cho việc lưu trữ, quản lý và trao đổi văn bản trên môi trường mạng.

Nhóm thực hiện dự thảo tiêu chuẩn khuyến nghị khi ban hành tên tiêu chuẩn nên là: “Thông tin và tư liệu - Tiêu chuẩn về định danh cho văn bản hành chính công điện tử dùng trong lưu trữ và trao đổi dữ liệu trên môi trường mạng” để thống nhất cách đặt tên với các tiêu chuẩn thuộc lĩnh vực "Thông tin và tư liệu" đã ban hành trước đó.