به نام خدا

بررسي محتوايي سايت Khamenei.ir – 28/11/1390

# ارتباط محتوايي

ارتباط داده‌هاي سايت را به دو گونه مي‌توان تقسيم كرد:

1. داده‌هايي كه از ابتدا با هم مرتبط مي‌باشند
2. داده‌هايي كه ارتباط اوليه ندارند

هنگامي كه فايل تصويري يك سخنراني به دفتر مي‌رسد، به فايل صوتي تبديل مي‌شود. فايل صوتي به صورت متني پياده شده و تبديل به متن بيانات مي‌گردد. عكس‌هاي گرفته شده از اين مراسم هم كه آمد، چهار نوع داده داريم كه همگي با هم مرتبط هستند. البته ارتباطي خام كه هنوز به صورت موضوعي تقطيع نشده‌اند.

اگر متن اين سخنراني تقطيع شده و برچسب‌گذاري شود، عملاً فيلم و صوت نيز تقطيع و برچسب‌دار شده است. به شرطي كه هنگام پياده نوار زمان‌بندي‌ها ثبت شود (Sync) و يا پس از تقطيع، قطعات به فايل صوتي منتقل شود (Resync). اگر فايل صوتي از فيلم استخراج شده باشد، تقطيع مزبور به خودي‌خود در فيلم نيز به دست خواهد آمد.

تا اين‌جا ما با يك‌بار برچسب‌گذاري (Tag)، سه نوع داده را تقطيع كرده‌ايم. امّا تصاوير به تمامي سخنراني به صورت يكپارچه متصل هستند و معنايي ندارد تقطيع شوند. طبيعتاً اخباري كه درباره اين سخنراني به دست مي‌رسند نيز مانند تصاوير، به صورت يك‌جا به سخنراني متصل هستند.

پس در اين نوع از داده‌هاي ورودي ما دو نوع خروجي خواهيم داشت:

1. داده‌هاي كلّي
   1. متن كامل سخنراني
   2. صوت كامل سخنراني
   3. فيلم كامل سخنراني
   4. اخبار مرتبط با سخنراني
   5. تصاوير مرتبط با سخنراني
2. داده‌هاي جزيي
   1. قطعات متن سخنراني
   2. قطعات صوتي سخنراني
   3. قطعات فيلم سخنراني

نوع ديگري از ورودي‌ها را هم داريم كه از ابتدا خبر يا نوشته مي‌باشند. پيام‌ها و انتصاب‌ها، تبريك‌ها و تسليت‌ها و تمامي اخباري كه از ديدارها و ملاقات‌ها و ساير فعاليت‌ها به دست مي‌رسد. اين داده‌ها قابليت تقطيع ندارند. پس در داده‌هاي كلّي قرار مي‌گيرند.

1. ادامه داده‌هاي كلّي
2. اخبار ديدارها، ملاقات و ساير فعاليت‌ها
3. پيام‌ها، انتصابات، تبريكات و تسليت‌ها

البته پاره‌اي از پيام‌ها در بعضي مناسبت‌هاي خاصّ ممكن است قابليت تقطيع داشته باشند كه به داده‌هاي جزيي تبديل خواهند شد، با حفظ وضعيت كلّي خود.

داده‌ها و اطلاعاتي نيز در خارج از دفتر نشر توليد مي‌شوند. اين اطلاعات همگي كلّي محسوب مي‌شوند و اهميت‌شان در حدّي نيست كه نياز به تقطيع داشته باشند.

به دليل دو نوع نگاه به اطلاعات؛ كلّي و جزيي، با دو نوع فرآيند برچسب‌زني مواجه خواهيم بود. اطلاعات جزيي هم‌زمان با تقطيع برچسب خواهند خورد و اطلاعات كلّي به دو صورت در دستور كار قرار خواهند گرفت؛

آن دسته از اطلاعات كلّي كه تقطيع شده‌اند نيازي به برچسب خوردن نخواهند داشت، زيرا پس از اتمام برچسب‌زني قطعات، يك فهرست غيرتكراري از برچسب‌ها به آن داده كلّي به صورت خودكار الصاق خواهد شد.

امّا پاره‌اي از اطلاعات كلّي، مانند: نوشته‌ها، اخبار، تصاوير و تمامي داده‌هايي كه از «ديگران» دريافت شده است، بايد برچسب‌زني شوند.

# بانك برچسب‌ها

هر برچسبي كه به داده‌اي ملصق مي‌شود، بايد در فهرستي قرار گيرد كه با تايپ يك حرف از ابتدا يا وسط، فراخواني شده و به نظر كارشناس برسد، تا كارشناس بتواند آن را انتخاب نمايد و نيازي به ايجاد برچسب جديد يا تكراري نداشته باشد.

# بانك مترادفات

مترادفات هر واژه يا اصطلاحي كه در برچسب‌ها ذكر شده در بانك مترادفات ذخيره خواهد شد. اين بانك هم در جستجوها ياري مي‌رساند و هم هنگام درج برچسب مي‌تواند كمك نمايد تا موارد تكراري به راحتي شناسايي شده و با اصطلاح اصلي جايگزين شوند. اصطلاح اصلي آن برچسبي است كه از ميان چند مترادف برگزيده مي‌شود تا به عنوان برچسب اوليه نمايش داده شده و به كار گرفته شود. واژه‌اي خواهد بود كه بيشتر كاربرد دارد و رايج‌تر است.

# گزينش برچسب‌هاي ويژه

آن‌چه امروز به عنوان واژه‌هاي اصلي «جستار» برگزيده مي‌شود، در حقيقت پاره‌اي از واژه‌هايي است كه به عنوان برچسب مورد استفاده قرار گرفته‌اند، ولي به دليل اهميت خاصّ، گزينش شده‌اند. در يك روند دائمي بايد چنين برچسب‌هايي شناسايي شوند و ارتباط‌شان با ساير برچسب‌ها مشخص شود. به اين ترتيب كه تمامي برچسب‌هايي كه ارتباط محتوايي با اين برچسب‌هاي منتخب دارند به آن‏ها متصل شوند.

# ايجاد پرونده

اين گزينش و ارتباط برچسب‌ها مقدمه‌اي است براي توليد «پرونده» براي داده‌ها و اطلاعات. پرونده مهم‌ترين بخش محتوايي سايت خواهد بود كه تمامي انواع داده را حول يك موضوع تجميع مي‌كند. پرونده در حقيقت توسعه‌يافته همان «جستار» مي‌باشد.

البته پرونده‌هاي ويژه بر جاي خواهند ماند؛ پرونده‌هايي كه موضوع بسيار خاصّ و ويژه دارند و در حقيقت گزينشي از پرونده‌هاي سايت خواهند بود كه به دليل اهميت فراوان به صورت پرونده ويژه معرفي خواهند شد.

# روش‌هاي دستيابي به برچسب

براي تسهيل ايجاد برچسب به غير از روش مرسوم، دو روش ديگر پيشنهاد مي‌گردد. در روش نخست، مي‌توان از متن بيانات به صورت رايانه‌اي و خودكار برچسب‌هاي اوليه را اسخراج كرد، به روش ذيل:

1. تمامي واژه‌ها تفكيك شده و در يك بانك اطلاعات هر كدام به صورت يك ركورد قرار مي‌گيرند
2. واژه‌هاي تكراري حذف شده و يك فهرست غيرتكراري در اختيار كارشناس قرار داده مي‌شود
3. غربال‌گر تمامي افعال، حروف و صفت‌ها و قيدهاي كم‌ارزش را حذف مي‌كند
4. واژه‌هاي باقي‌مانده مترادف‌يابي شده و موارد هم‌معنا به يكديگر متصل مي‌شوند
5. از گروه‌هاي مترادف، واژه اصلي انتخاب مي‌شود

اكنون حجم زيادي برچسب به دست خواهد آمد كه مي‌تواند پايه اصلي عمليات برچسب‌زني قرار بگيرد. خصوصاً با توجه به اين نكته كه تمامي اين واژگان از ادبيات رهبري استخراج شده‌اند.

روش ديگر مشابه همان است كه در «آرشيو پيشرفته‌» مورد استفاده قرار گرفته است؛ ايجاد رده‌هايي از اطلاعات پايه. بر اساس موضوعات نظام، اقشار، مناطق جغرافيايي و شهرها و…. مي‌توان رده‌هايي از اطلاعات پايه را در نظر گرفت و تمامي زيرمجموعه‌هاي آن را احصاء نمود و به عنوان بخشي از برچسب‌ها مورد استفاده قرار داد.

# يكپارچه‌سازي اطلاعات

اگر يك نظام واحد در برچسب‌زني اطلاعات به كار گرفته شود، به اين معني كه تمامي برچسب‌ها در يك استخر ريخته شده و از همان‌ اخذ شوند، ارتباط تمامي داده‌ها تضمين خواهد شد. خصوصاً با توجه به اين‏كه بعضي از اطلاعات از ابتدا با هم مرتبط بوده‌اند و در نظر گرفتن اين ارتباط اوليه، عمليات برچسب‌زني را ساده‌تر نموده است.

اين يكپارچگي از يك‌سو پرونده‌هاي موضعي قدرتمند و جامعي ايجاد مي‌كند كه در حقيقت هر برچسب ما را به يك پرونده مي‌رساند. از ديگرسو اطلاعات كلّي ديگران و خارج از دفتر را به اطلاعات داخلي متصل مي‌نمايد. اين ارتباط مي‌تواند بازتاب فرمايشات رهبري در رسانه‌ها و جامعه را قابل رهگيري نمايد. اين داده‌ها را مي‌توان در قالب بولتن‌هاي داخلي به استحضار مسئولان دفتر رساند.

# سرويس‌هاي موجود

فضاي فعلي سايت علي‌رغم حجم زياد داده‌ها و نظم بسيار خوبي كه دارند، اندكي سبب سردرگمي مي‌شود. يك كاربر معمولي در ابتداي ورود نمي‌داند بايد در بيانات جستجو كند يا سراغ آرشيو پيشرفته برود و يا در پرونده‌ها دنبال مطلب مورد نظر خود بگردد و شايد هم آن‌چه مي‌خواهد در بخش ديگران بهتر تبيين شده باشد.

اگر وحدت تمامي اطلاعات حول «پرونده» در نظر گرفته شود، در حقيقت «جستار» توسعه‏ يافته و محور تمام سايت مي‌شود. تمامي داده‌ها و اطلاعات حول پرونده قابل ملاحظه است.

يكي از مهم‌ترين استفاده كاربران نيز خبرخوان يا RSS است. اگر چه در حال حاضر در بخش‌هاي مختلف سايت اين گزينه قرار دارد، ولي جاي يك صفحه خالي است كه بتوان گزينه‌هاي مختلف را تيك زد و يه يك RSS دلخواه دست يافت كه تنها مثلاً مطالب مربوط به «اقتصاد» و شهر «قم» را نمايش دهد. بسياري از سايت‌هاي خبري امروزه از اين نوع RSSهاي قابل شخصي‌سازي استفاده مي‌نمايند.

اجركم عندا… ـ سيدمهدي موسوي موشَّح