



جمهوری اسلامی ایران

Islamic Republic of Iran

سازمان ملی استاندارد ایران

Iranian National Standardization Organization



استاندارد ملی ایران

۱۸۴۷۸

چاپ اول

۱۳۹۲

INSO

18478

1st. Edition

2014

فناوری اطلاعات – الگوریتم تطبیقی
بدون اتلاف فشرده‌سازی داده‌ها

**Information technology- Adaptive Lossless
Data Compression algorithm (ALDC)**

ICS : 35.040

به نام خدا

آشنایی با سازمان ملی استاندارد ایران

مؤسسه استاندارد و تحقیقات صنعتی ایران به موجب بند یک ماده ۳ قانون اصلاح قوانین و مقررات مؤسسه استاندارد و تحقیقات صنعتی ایران، مصوب بهمن ماه ۱۳۷۱ تنها مرجع رسمی کشور است که وظیفه تعیین، تدوین و نشر استانداردهای ملی (رسمی) ایران را به عهده دارد.

نام موسسه استاندارد و تحقیقات صنعتی ایران به موجب یکصد و پنجاه و دومین جلسه شورای عالی اداری مورخ ۹۰/۶/۲۹ به سازمان ملی استاندارد ایران تغییر و طی نامه شماره ۲۰۶/۳۵۸۳۸ مورخ ۹۰/۷/۲۴ جهت اجرا ابلاغ شده است.

تدوین استاندارد در حوزه های مختلف در کمیسیون های فنی مرکب از کارشناسان سازمان، صاحب نظران مراکز و مؤسسات علمی، پژوهشی، تولیدی و اقتصادی آگاه و مرتبط انجام می شود و کوششی همگام با مصالح ملی و با توجه به شرایط تولیدی، فناوری و تجاری است که از مشارکت آگاهانه و منصفانه صاحبان حق و نفع، شامل تولیدکنندگان، مصرف کنندگان، صادرکنندگان و وارد کنندگان، مراکز علمی و تخصصی، نهادهای سازمان های دولتی و غیر دولتی حاصل می شود. پیش نویس استانداردهای ملی ایران برای نظرخواهی به مراجع ذی نفع و اعضای کمیسیون های فنی مربوط ارسال می شود و پس از دریافت نظرها و پیشنهادات در کمیته ملی مرتبط با آن رشته، طرح و در صورت تصویب به عنوان استاندارد ملی (رسمی) ایران چاپ و منتشر می شود.

پیش نویس استانداردهایی که مؤسسات و سازمان های علاقه مند و ذی صلاح نیز با رعایت ضوابط تعیین شده تهیه می کنند در کمیته ملی طرح و بررسی و در صورت تصویب، به عنوان استاندارد ملی ایران چاپ و منتشر می شود. بدین ترتیب، استانداردهایی ملی تلقی می شوند که بر اساس مفاد نوشته شده در استاندارد ملی ایران شماره ۵ تدوین و در کمیته ملی استاندارد مربوط که سازمان ملی استاندارد ایران تشکیل می دهد به تصویب رسیده باشد.

سازمان ملی استاندارد ایران از اعضای اصلی سازمان بین المللی استاندارد (ISO)^۱، کمیسیون بین المللی الکتروتکنیک (IEC)^۲ و سازمان بین المللی اندازه شناسی قانونی (OIML)^۳ است و به عنوان تنها رابط^۴ کمیسیون کدکس غذایی (CAC)^۵ در کشور فعالیت می کند. در تدوین استانداردهای ملی ایران ضمن توجه به شرایط کلی و نیازمندی های خاص کشور، از آخرین پیشرفت های علمی، فنی و صنعتی جهان و استانداردهای بین المللی بهره گیری می شود.

سازمان ملی استاندارد ایران می تواند با رعایت موازین پیش بینی شده در قانون، برای حمایت از مصرف کنندگان، حفظ سلامت و ایمنی فردی و عمومی، حصول اطمینان از کیفیت محصولات و ملاحظات زیست محیطی و اقتصادی، اجرای بعضی از استانداردهای ملی ایران را برای محصولات تولیدی داخل کشور و/یا اقلام وارداتی، با تصویب شورای عالی استاندارد، اجباری نماید. سازمان می تواند به منظور حفظ بازارهای بین المللی برای محصولات کشور، اجرای استاندارد کالاهای صادراتی و درجه بندی آن را اجباری نماید. همچنین برای اطمینان بخشیدن به استفاده کنندگان از خدمات سازمان ها و مؤسسات فعال در زمینه مشاوره، آموزش، بازرسی، ممیزی و صدور گواهی سیستم های مدیریت کیفیت و مدیریت زیست محیطی، آزمایشگاه ها و مراکز کالیبراسیون (واسنجی) وسایل سنجش، سازمان ملی استاندارد ایران این گونه سازمان ها و مؤسسات را بر اساس ضوابط نظام تأیید صلاحیت ایران ارزیابی می کند و در صورت احراز شرایط لازم، گواهینامه تأیید صلاحیت به آنها اعطا و بر عملکرد آنها نظارت می کند. ترویج دستگاه بین المللی یکاها، کالیبراسیون (واسنجی) وسایل سنجش، تعیین عیار فلزات گرانبها و انجام تحقیقات کاربردی برای ارتقای سطح استانداردهای ملی ایران از دیگر وظایف این سازمان است.

1- International Organization for Standardization

2 - International Electrotechnical Commission

3- International Organization of Legal Metrology (Organisation Internationale de Metrologie Legale)

4 - Contact point

5 - Codex Alimentarius Commission

کمیسیون فنی تدوین استاندارد
« فناوری اطلاعات – الگوریتم تطبیقی بدون اتلاف فشرده سازی داده ها »

رئیس:

معاون فناوری ارتباطات مرکز تحقیقات صنایع انفورماتیک

صمدیان، علی
(لیسانس مهندسی الکترونیک)

دبیر:

سرپرست آزمایشگاه فناوری اطلاعات مرکز تحقیقات صنایع انفورماتیک

یحیایی، مهری
(فوق لیسانس مهندسی فناوری اطلاعات)

اعضا: (اسامی به ترتیب حروف الفبا)

کارشناس فنی مرکز تحقیقات صنایع انفورماتیک

آژ، رضوان
(لیسانس مهندسی کامپیوتر)

کارشناس فنی مرکز تحقیقات صنایع انفورماتیک

تورانی، فرزاد
(لیسانس مهندسی کامپیوتر)

کارشناس شرکت ارتباطات زیرساخت

زندباف، عباس
(لیسانس مهندسی الکترونیک-مخابرات)

کارشناس فنی مرکز تحقیقات صنایع انفورماتیک

شاهی، فرید
(لیسانس مهندسی کامپیوتر)

کارشناس استاندارد سازمان تنظیم مقررات و ارتباطات رادیویی

عروجی، سیدمهدی
(فوق لیسانس مدیریت فناوری اطلاعات)

عضو هیات علمی دانشگاه علم و صنعت

نادری، مجید
(دکترای مهندسی برق - الکترونیک)

فهرست مندرجات

صفحه	عنوان
ب	آشنایی با سازمان ملی استاندارد ایران
ج	کمیسیون فنی تدوین استاندارد
و	پیش‌گفتار
ز	مقدمه
۱	۱ هدف و دامنه کاربرد
۱	۲ انطباق
۱	۳ مراجع الزامی
۱	۴ اصطلاحات و تعاریف
۲	۴-۱ جریان داده فشرده‌سازی شده
۲	۴-۲ اشاره‌گر رونوشت
۲	۴-۳ آدرس کنونی
۲	۴-۴ بایت داده
۲	۴-۵ فیلد جابجایی
۲	۴-۶ نشانگر پایان:
۲	۴-۷ بافر تاریخچه
۲	۴-۸ نویسه
۲	۴-۹ رشته همسان
۲	۴-۱۰ تعداد تطابق
۲	۴-۱۱ فیلد تعداد تطابق
۳	۴-۱۲ بیت‌های پُرکننده
۳	۵ قراردادهای و نشان‌گذاری‌ها
۳	۵-۱ نماینده اعداد
۳	۵-۲ نامها
۳	۶ الگوریتم فشرده‌سازی ALDC
۳	۶-۱ توصیف کُدگذاری بافر تاریخچه ۵۱۲ بیتی
۵	۶-۲ توصیف جریان داده فشرده شده

فهرست مندرجات-ادامه

صفحه	عنوان
۷	پیوست الف (اطلاعاتی) قالب کدگذاری ALDC
۱۰	پیوست ب (اطلاعاتی) مرور کلی ALDC
۱۱	پیوست پ (اطلاعاتی) نمودار جریان کدگذاری ALDC
۱۵	پیوست ت (اطلاعاتی) کتابنامه

پیش‌گفتار

استاندارد «فناوری اطلاعات – الگوریتم تطبیقی بدون اتلاف فشرده‌سازی داده‌ها» که پیش‌نویس آن در کمیسیون‌های مربوط، توسط مرکز تحقیقات صنایع انفورماتیک به عنوان استاندارد ملی ایران، تهیه و تدوین شده و در دویست و نود و ششمین اجلاس کمیته ملی استاندارد رایانه و فناوری داده‌ها مورخ ۹۲/۹/۱۹ مورد تصویب قرار گرفته است، اینک به استناد بند یک ماده ۳ قانون اصلاح قوانین و مقررات سازمان ملی استاندارد ایران، مصوب بهمن ماه ۱۳۷۱، به عنوان استاندارد ملی ایران منتشر می‌شود.

برای حفظ همگامی و هماهنگی با تحولات و پیشرفت‌های ملی و جهانی در زمینه صنایع، علوم و خدمات، استانداردهای ملی ایران در مواقع لزوم تجدید نظر خواهد شد و هر پیشنهادی که برای اصلاح و تکمیل این استانداردها ارائه شود، هنگام تجدید نظر در کمیسیون فنی مربوط مورد توجه قرار خواهد گرفت. بنابراین، باید همواره از آخرین تجدید نظر استانداردهای ملی استفاده کرد.

منبع و مآخذی که برای تهیه این استاندارد بکار گرفته شده، به شرح زیر است:

ISO/IEC 15200:1996, Information technology — Adaptive Lossless Data Compression algorithm (ALDC)

مقدمه

در دهه‌های گذشته، سازمان ISO/IEC استانداردهای بین‌المللی فراوانی را دربارهٔ نوارهای مغناطیسی^۱، کاست‌های نوار مغناطیسی و کارتریج‌های مغناطیسی و کارتریج‌های دیسک نوری چاپ کرده است. وسایل توسعه یافته اخیر، تراکم ذخیره‌سازی فیزیکی^۲ خیلی بالایی دارند. برای بهره‌برداری بهینه در استفاده از این ظرفیت داده بدست آمده، الگوریتم‌های فشرده‌سازی بدون اتلاف طراحی شده‌اند، که این الگوریتم‌ها امکان کاهش تعداد بیت‌های لازم برای نمایش داده‌های کاربران را فراهم می‌آورند.

این الگوریتم‌های فشرده‌سازی، توسط ECMA (که به عنوان مرجع صلاحیت‌دار بین‌المللی ثبت^۳ شده توسط ISO/IEC مشخص شده) ثبت شده‌اند. این ثبت شامل تخصیص یک شناساگر عددی^۴ به هر الگوریتم ثبت شده می‌باشد. این شناساگر عددی در رسانه^۵ ذخیره می‌گردد و نشان می‌دهد که چه الگوریتم (یا الگوریتم‌های) فشرده‌سازی استفاده شده‌اند.

این استاندارد ملی ایران، سومین استاندارد از نوع خود برای الگوریتم‌های فشرده‌سازی بدون اتلاف است. دو استاندارد قبلی در این مورد عبارتند از:

استاندارد ملی ایران شماره ۱۲۰۴۲: سال ۱۳۹۰، فناوری اطلاعات - فشرده‌سازی داده‌ها برای تبادل اطلاعات - الگوریتم کدگذاری حساب دودویی

ISO/IEC 11558:1992, Information technology — Data compression for information interchange — Adaptive coding with embedded dictionary — DCLZ algorithm.

1- Magnetic tapes
2- Physical recording density
3- International Registration Authority
4- Numerical identifier
5- Medium

فناوری اطلاعات – الگوریتم تطبیقی بدون اتلاف فشرده سازی داده‌ها

۱ هدف و دامنه کاربرد

هدف از تدوین این استاندارد ملی تعریفی برای الگوریتم بدون اتلاف فشرده‌سازی جهت کاهش تعداد بایت‌های لازم برای نمایش داده می‌باشد. این الگوریتم، با عنوان الگوریتم تطبیقی بدون اتلاف فشرده‌سازی داده^۱ نامیده می‌شود. شناساگرهای عددی اختصاص داده شده به این الگوریتم، مطابق استاندارد ملی ایران به شماره ۱۱۵۷۶، بدین شرح است:

بافر تاریخچه ۵۱۲ بیتی ALDC^۲ : ۳

بافر تاریخچه ۱۰۲۴ بیتی ALDC : ۴

بافر تاریخچه ۲۰۴۸ بیتی ALDC : ۵

۲ انطباق

هر الگوریتم فشرده‌سازی در صورتی با این استاندارد ملی مطابق است که جریان داده خروجی^۳ حاصل از آن، الزامات این استاندارد ملی ایران را برآورده سازد.

۳ مراجع الزامی

مدارک الزامی زیر حاوی مقرراتی است که در متن این استاندارد ملی ایران به آنها ارجاع داده شده است. بدین ترتیب آن مقررات، جزئی از این استاندارد ملی ایران محسوب می‌شود. در صورتی که به مدرکی با ذکر تاریخ انتشار ارجاع داده شده باشد، اصلاحیه‌ها و تجدید نظرهای بعدی آن، مورد نظر این استاندارد ملی ایران نیست. در مورد مدارکی که بدون ذکر تاریخ انتشار به آنها ارجاع داده شده است، همواره آخرین تجدید نظر و اصلاحیه‌های بعدی آنها مورد نظر است. استفاده از مراجع زیر، به اضافه مراجع ذکر شده در این بند، برای این استاندارد الزامی است: استاندارد ملی ایران به شماره ۱۱۵۷۶: سال ۱۳۸۹، فناوری اطلاعات - روالی برای ثبت الگوریتم‌های فشرده‌سازی بدون از دست دادن داده.

۴ اصطلاحات و تعاریف

برای مقاصد این استاندارد ملی ایران، تعاریف زیر به کار می‌روند:

1- ALDC
2 - ALDC 512-Byte History Buffer
3 - Output data stream

- ۱-۴ جریان داده فشرده‌سازی شده^۱: جریان خروجی بعد از کُدگذاری^۲ می‌باشد.
- ۲-۴ اشاره‌گر رونوشت^۳: بخشی از جریان داده فشرده شده می‌باشد که نمایش‌دهنده گروهی از دو یا چند بایت متوالی است و برای این گروه، گروهی مشخص در بافر تاریخچه موجود است. این شامل فیلد طول کُد^۴ و فیلد جابجایی^۵ است.
- ۳-۴ آدرس کنونی: محلی در بافر تاریخچه که بایت داده^۶ در آنجا نوشته شده است.
- ۴-۴ بایت داده: بایت کنونی داده ورودی که در بافر تاریخچه نوشته شده است و با تمام بایت‌های داده‌ای که از قبل در بافر تاریخچه نوشته شده‌اند، مقایسه می‌شود.
- ۵-۴ فیلد جابجایی: قسمتی از اشاره‌گر رونوشت که محلی را در بافر تاریخچه اولین بایت یک رشته همسان^۷ مشخص می‌نماید.
- ۶-۴ نشانگر پایان^۸: رشته‌ای از ۱۲ عدد یک منطقی^۹ است و پایان یک جریان داده فشرده‌سازی شده را مشخص می‌نماید.
- ۷-۴ بافر تاریخچه: ساختمان داده‌ای^{۱۰} که بایت‌های داده ورودی برای استفاده در فرآیندهای فشرده‌سازی و گشوده‌سازی^{۱۱}، در آن ذخیره می‌شوند.
- ۸-۴ نویسه^{۱۲}: بایت داده‌ای که هیچ همسانی با آن در بافر تاریخچه، یافت نشود.
- ۹-۴ رشته همسان: دنباله‌ای از بایت‌ها در داده‌های ورودی که با دنباله‌ای از بایت‌ها در بافر تاریخچه، مشخص شده است.
- ۱۰-۴ تعداد تطابق^{۱۳}: تعداد بایت‌های موجود در یک رشته همسان می‌باشد.
- ۱۱-۴ فیلد تعداد تطابق: قسمتی از اشاره‌گر رونوشت که مشخص‌کننده تعداد بایت‌های پیاپی است که برای آنها یک همسان در بافر تاریخچه یافت شده است.

-
- 1 - Compressed Data Stream
 - 2 - Encoding
 - 3 - Copy Pointer
 - 4 - Length Code Field
 - 5 - Displacement Field
 - 6 - Data Byte
 - 7 - Matching String
 - 8 - End Marker
 - 9 - ONE
 - 10- Data structure
 - 11- Decompression
 - 12- Literal
 - 13- Match Count

۴-۱۲ بیت‌های پُرکننده^۱: بیت‌هایی که نیاز است برای حفظ^۲ مرز ۸ بیتی بودن یک بایت، بر روی صفر منطقی^۳ تنظیم شده و در جریان داده فشرده شده، قرار داده شوند.

۵ قراردادها و نشان‌گذاری‌ها^۴

۵-۱ نماینده اعداد

قراردادها و نشان‌گذاری‌های زیر در این استاندارد ملی ایران معتبر هستند، مگر غیر از آن ذکر شود.

- تنظیمات بیت‌های دودویی با صفر منطقی و یک منطقی نشان داده می‌شود.
- اعداد نشان‌گذاری شده به صورت دودویی و ترکیب‌های بیت با صفرها و یک‌های منطقی توسط با ارزش‌ترین بیت^۵ در سمت چپ با صفرها و یک‌ها نمایش داده شده‌اند.
- همه اعداد دیگر باید به صورت دهدهی^۶ باشند.

۵-۲ نام‌ها

نام‌های هستارها^۷ با یک حرف ایرانیک، نمایش داده شده است.

۶ الگوریتم فشرده‌سازی ALDC

چکیده‌ای از فرآیند کُدگذاری ALDC در پیوست ب، و نمودار جریان^۸ آن در پیوست پ آمده‌اند.

۶-۱ توصیف کُدگذاری بافر تاریخچه ۵۱۲ بیتی

در شروع کُدگذاری، تمام بایت‌ها در بافر تاریخچه باید به صفرهای منطقی بازنشانده شوند^۹. بایت‌های داده باید به ترتیب در بافر تاریخچه ذخیره شوند، آدرس کنونی باید با ۰ شروع شود.

در هر زمان، کُدگذار^{۱۰} یک بایت از جریان داده ورودی را پردازش می‌کند. به بایتی که هم اکنون در حال پردازش شدن است، با عنوان بایت داده ارجاع می‌شود. وقتی بایت داده از جریان داده ورودی دریافت می‌شود، باید جایی که آدرس کنونی در بافر تاریخچه اشاره می‌کند، نوشته شود. سپس آدرس کنونی باید به ۱ افزایش یابد. اگر مقدار آن از بیشینه آدرس فراتر رود، که در مورد اندازه بافر تاریخچه ۵۱۲ بیتی این مقدار ۵۱۱ است، باید با مقدار ۰ بازنشانی شود.

1 - Pad Bits
2 - To maintain
3 - ZERO
4 - Conventions and Notations
5 - Most significant bit
6 - Decimal form
7 - Entities
8 - Flow chart
9 - Reset
10- Encoder

گام ۱

بایت داده باید با هر یک از بایت‌هایی که از قبل در بافر تاریخچه نوشته شده‌اند، مقایسه شوند تا همه بایت‌های یکسان^۱، شناسایی شوند.

گام ۲

- اگر بایت داده با هیچ یک از بایت‌های بافر تاریخچه یکسان نبود، فرآیند باید در گام ۶ ادامه پیدا کند.
- اگر بایت داده با یک یا چند بایت در بافر تاریخچه همسان بود، باید برای هر بایت همسان مشخص گردد که آیا این بایت همسان در ادامه دنباله قبلی از بایت‌های همسان است یا خیر.
 - اگر در ادامه چنین دنباله‌ای نبود، فیلد جابجایی بایت همسان باید مشخص و ثبت گردد که تعداد انطباق به اندازه یک بایت دارد.
 - اگر بایت همسان در ادامه رشته قبلی بود، تعداد انطباق برای همان رشته باید به ۱ افزایش یابد.

گام ۳

- اگر تعداد انطباق با ۲۷۱ برابر شد، بایت‌های مربوط باید با یک اشاره‌گر رونوشت، که به جریان داده فشرده سازی شده اضافه می‌شوند، مشخص گردند. فیلد شمار همسانی و فیلد جابجایی آن باید مطابق آنچه در زیربند ۶-۲ آمده است، تعیین شوند. بایت داده بعدی باید خوانده شود و فرآیند از گام ۱ ادامه یابد.
- اگر داده دیگری برای خواندن وجود نداشت، فرآیند باید از گام ۷ ادامه یابد.
- یادآوری - مقدار ۲۷۱ به دلایل پیاده‌سازی انتخاب شده است.

گام ۴

- اگر شمار همسانی به مقدار ۲۷۱ نرسیده است، هر یک از رشته‌های همسان معوق^۲، برای فهمیدن اینکه بایت داده در ادامه آنها آمده است، باید بررسی شود.
- اگر هیچ یک از رشته‌های همسان قبلی ادامه نیافته باشد و اگر هیچ یک از رشته‌های همسان قبلی، از دو یا چند بایت تشکیل نشده باشند، آن رشته همسانی که کمترین فیلد جابجایی را دارد باید بوسیله یک اشاره‌گر رونوشت تعیین گردد، همین اشاره‌گر رونوشت باید به جریان داده فشرده‌سازی شده اضافه شود. سپس بایت داده بعدی باید خوانده شود. اگر داده بیشتری برای خواندن وجود نداشت، فرآیند باید از گام ۷ ادامه یابد.
 - اگر هیچ یک از رشته‌های همسان قبلی ادامه نیافته باشند و همسان‌های قبلی نیز فقط یک بایت همسان باشند، همسان یک بایتی قبلی باید به عنوان نویسه تعیین گردد و همان‌گونه که در زیربند ۶-۲ آمده است، باید به جریان داده فشرده‌سازی شده اضافه گردد. سپس باید بایت داده بعدی خوانده شود. اگر داده دیگری برای خواندن وجود نداشت، فرآیند باید از گام ۷ ادامه یابد.

1- Identical bytes

2 -Pending Matching Strings

گام ۵

اگر هیچ رشته همسانی وجود نداشت که شمار همسان آن با ۲۷۱ برابر شده باشد و نیز ادامه‌ای از دست‌کم یک رشته همسان قبلی وجود داشت، بایت داده بعدی باید خوانده شود. فرآیند باید از گام ۱ ادامه یابد.

اگر داده بیشتری برای خواندن وجود نداشت، رشته همسان معلق باید به عنوان یک اشاره‌گر رونوشت تعیین شود که باید به جریان داده فشرده‌سازی شده اضافه گردد. فرآیند باید از گام ۷ ادامه یابد.

گام ۶

اگر بایت داده با هیچ یک از بایت‌های بافر تاریخچه همسان نبود، همه رشته‌های همسان قبلی که ممکن است معلق باشند باید بررسی شوند.

اگر هر رشته همسانی با دو یا چند بایت از قبل وجود داشت، باید آنها را بوسیله یک اشاره‌گر رونوشت، که به جریان داده فشرده‌سازی شده اضافه می‌شود، مشخص نمود. سپس بایت داده‌ای که همسانی ندارد باید به عنوان یک نویسه به جریان داده فشرده شده اضافه شود. پس از آن بایت داده بعدی باید خوانده شود و فرآیند باید از گام ۱ ادامه یابد. اگر داده اضافی برای خواندن وجود نداشت، فرآیند باید از گام ۷ ادامه یابد.

اگر هیچ گونه رشته‌های همسان پیوسان شامل دو یا چند بایت و نیز هیچ همسان یک بایتی معلق وجود نداشت، بایت قبل از بایت داده، باید به عنوان یک نویسه تعیین گردد و این نویسه به جریان داده فشرده‌سازی شده باید اضافه شود. بایت داده بعدی باید خوانده شود و فرآیند باید از گام ۱ ادامه یابد. اگر داده دیگری برای خواندن وجود نداشت، فرآیند باید از گام ۷ ادامه یابد.

گام ۷

یک نشانگر پایان باید به جریان داده فشرده اضافه می‌شود و در صورت نیاز بیت‌های پُرکننده را باید شامل شوند. فرآیند کدگذاری در این گام پایان می‌یابد.

۶-۲ توصیف جریان داده فشرده شده

همانگونه که در بالا توضیح داده شد، پردازش داده‌های ورودی جریان داده فشرده‌سازی شده را، به عنوان خروجی خود تولید می‌کند. جریان داده فشرده تکمیل شده باید شامل این موارد باشد:

- نویسه‌هایی که پس از هر کدام آنها، یک بیت تنظیم شده با صفر منطقی آمده است.
- اشاره‌گرهای رونوشت که پس از هر کدام آنها، یک بیت تنظیم شده با یک منطقی آمده است.
- یک نشانگر پایان که پس از آن، یک بیت تنظیم شده با یک منطقی آمده است.
- بیت‌های پُرکننده.

یکبار که تمام داده‌ها خوانده شدند، جریان داده فشرده‌سازی شده بوسیله یک بیت تنظیم شده با یک منطقی، که پس از آن نشانگر پایان و سپس بیت‌های پُرکننده آمده است، باید خاتمه یابد.

اگر بیش از یک رشته همسان از تعداد انطباق‌های یکسان در حین کدگذاری یافت شد، باید از آن اشاره‌گر رونوشت که کمترین فیلد جابجایی را دارد، استفاده شود. فیلد شمار همسانی باید شامل ۲، ۴، ۶، ۸، یا ۱۲ بیت باشد که شمار همسانی‌ها را همانگونه که در جدول ۱ مشخص شده است، تعیین نمایند. طول فیلد

جابجایی برای اندازه‌های بافرهای تاریخچه ۵۱۲ بایتی، ۱۰۲۴ بایتی، یا ۲۰۴۸ بایتی به ترتیب باید ۹ بیت، ۱۰ بیت، یا ۱۱ بیت باشد.

جدول ۱- فیلدِ شمارِ همسانی

تعداد بایت‌ها	تنظیم فیلدِ شمارِ همسانی
۲	00
۳	01
۴	10 00
:	:
:	:
۷	10 11
۸	110 000
:	:
:	:
۱۵	110 111
۱۶	1110 0000
:	:
:	:
۳۱	1110 1111
۳۲	1111 0000 0000
:	:
:	:
۲۷۰	1111 1110 1110
۲۷۱	1111 1110 1111

پیوست الف

(اطلاعاتی)

قالب^۱ کدگذاری ALDC

الف-۱ فهرست نمادها^۲

قالب کدگذاری جریان داده فشرده با استفاده از زبان شبیه به BNF^۳ توصیف شده است. نمادها^۴ در زیر تعریف شده‌اند:

نماد	تعریف
= :	عبارت غیر پایانی ^۵ سمت چپ "=" می‌تواند با عبارت سمت راست جایگزین شود.
<name>	عبارت غیر پایانی
[]	عبارت درون کروشه‌ها مجاز است یک یا چندین بار ظاهر شود.
	انفصال منطقی یا ^۶
()	متن قرار گرفته در داخل کمان‌ها ^۷ فقط توضیح هستند و برای روشن شدن موضوع اضافه شده‌اند.
0,1	اعداد دودویی پایانی ^۸ که صفر منطقی و یا یک منطقی هستند.

-
- 1 - Format
 - 2 - Nomenclature
 - 3 - Backus Normal Form
 - 4 - Symbols
 - 5 - Non-terminal
 - 6 - OR
 - 7 - Parentheses
 - 8 - Terminal

الگوریتم ALDC_1 - قالب کدگذاری (بافر تاریخچه ۵۱۲ بایتی)

< Match Count Field > := (مے تواند طولی، برابر ۲، ۴، ۶، ۸ یا ۱۲ بیت داشته باشد)

تنظیم فیلدِ شمارِ همسانی	تعداد بیت‌های شمارِ همسانی
00	۲
01	۳
10 00	۴
:	:
:	:
10 11	۷
110 000	۸
:	:
:	:
110 111	۱۵
1110 0000	۱۶
:	:
:	:
1110 1111	۳۱
1111 0000 0000	۳۲
:	:
:	:
1111 1110 1110	۲۷۰
1111 1110 1111	۲۷۱

تعداد بیت‌هایی که بر روی صفر منطقی تنظیم می‌شوند به نگهداری مرز ۸ بیتی نیاز دارند < Pad > := 1111 1111 1111
< End Marker > := 1111 1111 1111

< Displacement Field > := (۱۰ بیت)

الف-۴ الگوریتم ALDC_4 - قالب کدگذاری (بافرِ تاریخچه ۲۰۴۸ بایتی)

(به جز در مورد فیلدِ جابجایی که ۱۱ بیتی است، تعریف آن با ALDC_1 یکسان است)

(۱۱ بیت) < Displacement Field > :=

پیوست ب (اطلاعاتی)

مرور کلی ALDC

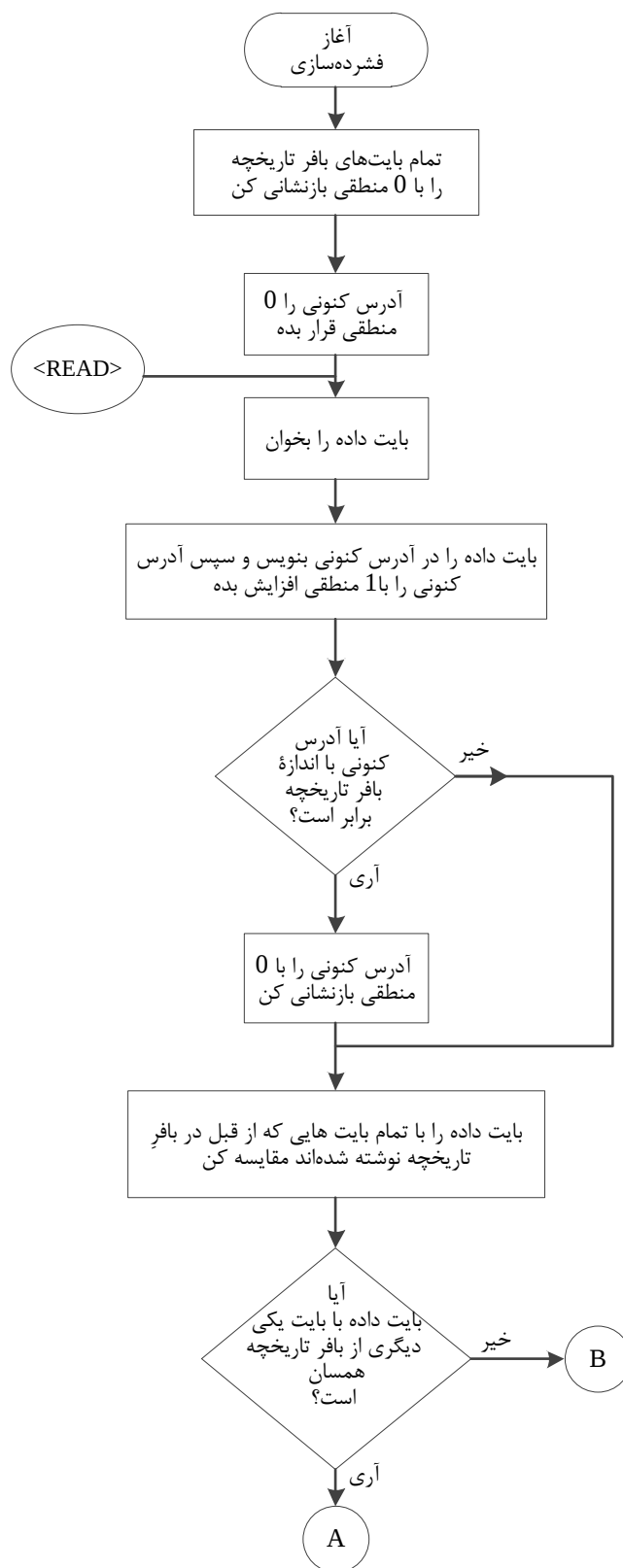
الگوریتم ALDC، ورودی را به صورت بایت‌های داده ۸ بیتی می‌پذیرد و جریانی از بیت‌ها را که نمایش دهنده داده‌ها به شکل فشرده شده هستند را در خروجی می‌گذارد. الگوریتم ALDC یک پیاده‌سازی از الگوریتم‌های فشرده‌سازی داده‌ی رده LZ1^۱ می‌باشد.

الگوریتم‌های LZ1 با استفاده از ساختمان داده‌ای که بافر تاریخچه نامیده می‌شود، فشرده‌سازی را محقق می‌کند. در بافر تاریخچه، داده ورودی ذخیره می‌گردد و با داده‌های پیشین در همان بافر تاریخچه مقایسه می‌شود. هر دو فرآیند کدگذاری LZ1 و فرآیند کُدگشایی^۲ LZ1، این ساختمان داده را از نو به حالت معین یکسانی برمی‌گردانند و با روشی مشابه آنرا به‌روز رسانی می‌کنند. به همین دلیل، این دو تاریخچه یکسان باقی می‌مانند و نیازی نخواهد بود اطلاعات محتویات تاریخچه در جریان داده فشرده قرار گیرد.

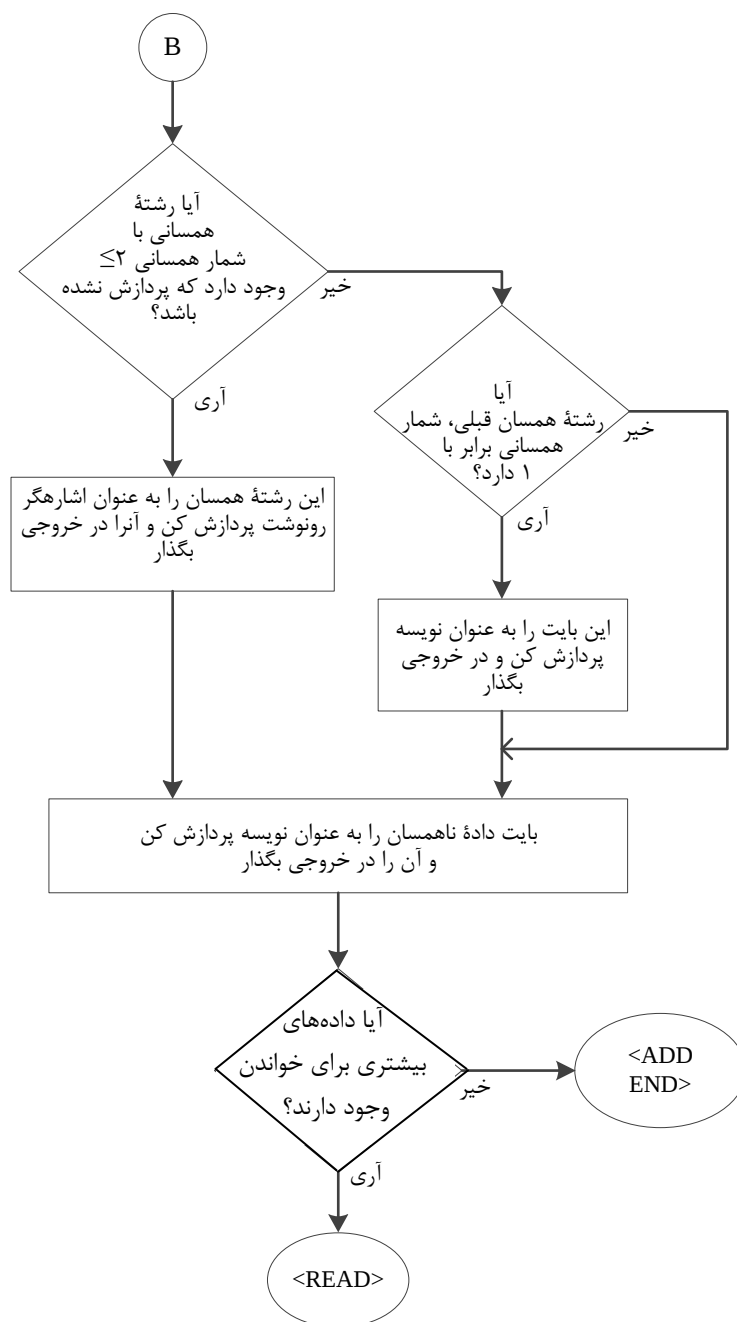
داده‌های ورودی به بافر تاریخچه وارد می‌شوند. هر بایت ورودی با تمام بایت‌هایی که از قبل در بافر تاریخچه ذخیره شده‌اند، مقایسه می‌شود. نتایج فشرده‌سازی از پیدا کردن همسان‌های متوالی ناشی می‌شود. در آغاز فرآیند کدگذاری، بافر تاریخچه خالی است. در صورت پیدا نشدن هیچ بایت همسانی در بافر تاریخچه، آنگاه جریان داده فشرده‌سازی شده فقط شامل نویسه‌ها خواهد بود. بایت‌ها تا زمانی به عنوان نویسه، کدگذاری می‌شوند که دنباله‌ای همسان از دو یا چند بایت ظاهر شود. هنگامی که بافر تاریخچه پُر می‌شود، بطور افزاینده‌ای برای کُدگذار این امکان فراهم می‌شود تا داده‌های ورودی را به نمایش بگذارد، این کار با استفاده از کدگذاری داده‌های ورودی به عنوان یک اشاره‌گر رونوشت برای رشته‌ای که از قبل در این بافر تاریخچه موجود بوده است، انجام می‌شود. این سازوکار اصلی است که الگوریتم‌های LZ1 با استفاده از آن می‌توانند به فشرده‌سازی دست یابند.

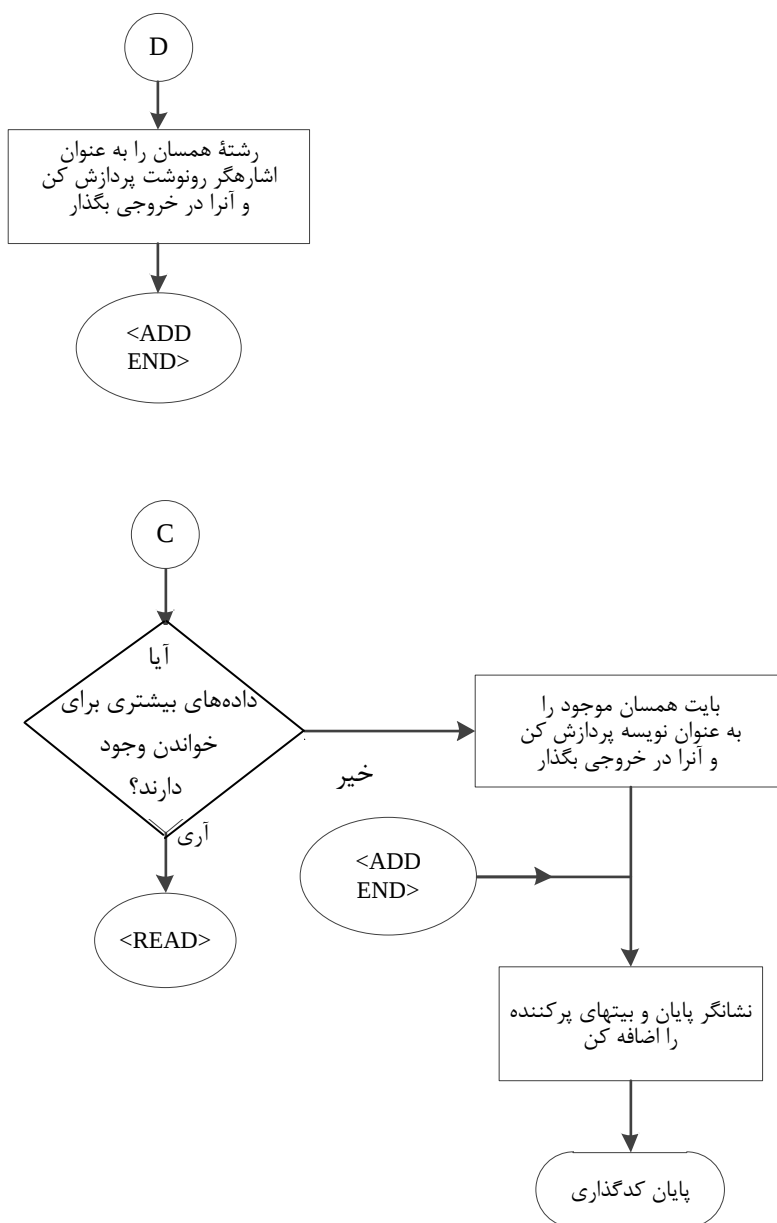
1 - Lempel-Ziv
2 - Decoding

پیوست پ
(اطلاعاتی)
نمودار جریانِ کدگذاری ALDC









پیوست ت
(اطلاعاتی)
کتابنامه

J. Ziv and A. Lempel, "A universal algorithm for sequential data compression" IEEE Trans. Inform. Theory, vol. IT-23, no.3, pp.337-343, 1977.